Beyond face value: involuntary emotional anticipation in typical development and Asperger's syndrome

being a Thesis submitted for the Degree of

Doctor of Philosophy

University of Hull

Letizia Palumbo

MSc in Psychology

"La Sapienza" University of Rome

March 2012

To my grandparents

Abstract

Understanding and anticipating the behavior and associated mental/emotional states of mind of others is crucial for successful social interactions. Typically developed (TD) humans rely on the processing and integration of social cues that accompany other's actions to make, either implicitly or explicitly, inferences about others' mental states. Interestingly, the attribution of affective or mental states to the agent can in turn (top down) induce distortions in the visual perception of those actions (Hudson, Liu, & Jellema, 2009; Hudson & Jellema, 2011; Jellema, Pecchinenda, Palumbo, & Tan, 2011). The aim of this thesis was to investigate bottom-up and top-down influences on distortions in the perception of dynamic facial expressions and to explore the role those biases may play in action/emotion understanding.

In Chapter 2 a series of six experiments examined the roles played by basic perceptual processes such as contrast effects (Thayer, 1980; Russell & Fehr, 1987; Tanaka-Matsumi, Attivissimo, Nelson, & D'Urso, 1995; Suzuki & Cavanagh, 1998), adaptation (Fox & Barton, 2007; Hsu & Young, 2004; Rutherford, Chattha, & Krysko, 2008; Webster Kaping, Mizokami, & Duhamel, 2004) and extrapolation or representational momentum (RM; Freyd & Finke, 1984; Freyd, 1987; Yoshikawa & Sato, 2008) and by 'emotional anticipation', i.e. the involuntary anticipation of the other's emotional state based on the immediate perceptual history.

In Experiment 1 video-clips were used showing a facial expression of joy or anger of which the emotional intensity gradually decreased until a neutral expression (or a 10% joy or anger expression) was reached. Participants' task was to evaluate the last neutral frame of the video sequence using a 5-point scale

ranging from slightly angry via neutral to slightly happy. The results showed that the expression depicted in the last frame of the happy-to-neutral videos was judged as slightly angry, while the same neutral expression depicted at the end of the angry-to-neutral videos was judged as slightly happy (‘overshoot’ response bias).

In Experiment 2 the identity of the actors was changed in the last neutral frame. Therefore, the observer was requested to make a judgment about an identity that was new to the observer; i.e. someone for whom no immediate perceptual history was available. If the overshoot bias would reflect sequential contrast/context effects or RM on an underlying valence dimension, then a change in identity should not affect the bias. On the other hand, the emotional anticipation hypothesis would predict the absence of an overshoot bias as the observer cannot anticipate the emotional state of someone on the basis of a perceptual history that pertains to a different identity. The Identity change manipulation removed the overshoot, suggesting that neither sequential contrast/context effects nor RM played a major role. However, the results can be explained by the emotional anticipation account. The findings of Experiment 2 may also suggest that adaptation does not play a crucial role as aftereffects are transfered to different identity, although to a less extent.

In Experiment 3 the starting expression of the video-clips was a neutral expression, which morphed via happy (or angry) back to the same neutral expression (joy or anger ‘loops’). In this ‘loop’ condition the contrast between the initial and last frames was absent as both showed identical neutral expressions. Results showed a response bias, which was opposite in direction (‘undershoot’ bias) to the bias in the control condition. These findings are not compatible with a sequential contrast hypothesis, as this hypothesis did not predict the presence of a bias, nor are they compatible with RM on an underlying positive-negative valence dimension, as in this case an overshoot bias was expected. However, the ‘undershoot’ effect may reflect an implicit attribution about the actor’s emotional disposition towards the observer based on the immediate perceptual history (positive disposition in the joy-loop and a negative disposition in the anger-loop).

In Experiment 4 explanations based on RM effects, which effects are known to be positively correlated with the speed of the intensity change, were excluded as variations of the speed of the video-clips did not affect the overshoot bias.

In Experiment 5 a mask (a textured object) was inserted to interrupt the flow of the sequence for 400 ms immediately before the last, neutral expression. The overshoot bias occurred also in this condition, suggesting that it did not depend on the uninterrupted flow of motion leading up to the last neutral frame of the video-sequence.

Finally, visual adaptation was further ruled out in Experiment 6 as the overshoot was not affected by doubling the duration of the first frame from 300 to 600 ms.

Overall, the findings reported in Chapter 2 suggested that the overshoot bias cannot be explained by basic bottom-up visual processes. Alternatively, the direction of the bias is congruent with the emotional anticipation account, which would explain the overshoot as the consequence of top-down attributions made by the observer about what the emotional state of the agent would be in the immediate future, on the basis of the immediate perceptual history.

In Chapter 3 four experiments focused to test the ‘emotional anticipation’ task on individuals with Asperger’s syndrome (AS). These individuals show impairments in communicative and social domains as well as limited empathy. They have been reported to fail to mentalize spontaneously and they are more inclined to understand human actions in a systemizing manner, representing them in terms of physical characteristics and forces. Individuals with AS were tested on the emotional anticipation tasks to see whether they would anticipate changes in the other's emotional state from the dynamic expressivity of the face, or whether they would not possess this emotional anticipation ability. In the latter case it might be that they would try and compensate for the lack of emotional anticipation.

Experiment 7, consisting of short video-clips of facial expressions (100% joy or 100% anger) that gradually morphed into a (nearly) neutral expression, revealed that AS participants showed an overshoot bias in their judgments of the final expression of the videos, which was very similar to that of the TD individuals. At first sight this might suggest that the AS individuals, just like the TD individuals, used emotional anticipation to solve the task. However, when applying the 'Identity change' manipulation in Experiment 8, the response bias was absent in the TD group, but remained present in the AS group. This finding argued against the AS individuals possessing an involuntary emotional anticipation ability. However, if they do not possess this ability, then what did cause them to show the overshoot bias in Experiment 7?

In Experiment 9, the overshoot response bias of the AS individuals was not affected by the Mask manipulation, similarly to the TD group. This suggests that direct extrapolation of the geometries from dynamic facial expressions did not play a major role in bringing about the perceptual distortion in the AS individuals.

In Experiment 10, the 'Loop' manipulation, involving the Neutral-to-joy-to-neutral and Neutral-to-anger-to-neutral sequences, showed that AS and TD groups again diverted, with AS individuals judging the last neutral frame as nearly neutral, whereas the TD showed an 'undershoot' bias. This suggested that the AS individuals could have relied on sequential contrast effects in their evaluations of the expressions.

On the whole, these findings suggest that in AS individuals the perceptual judgments of other's facial expressions are not influenced by emotional anticipation. The results suggest that they compensated for their failure to spontaneously attribute/anticipate emotional states to others by relying on sequential contrast effects.

In Chapter 4 the emotional anticipation hypothesis was further tested in TD individuals using deceptive, or 'fake', facial expressions in the Joy-to-neutral and Anger-to-neutral video-clips, displayed by professional actors. The rationale for using faked expressions was that as faked expressions reflect an emotional state that is incongruent with the literal facial features displayed, the emotional anticipation hypothesis would predict to find no, or a reduced, overshoot response bias compared to genuine dynamic expressions.

In Experiment 11 the typical overshoot response bias was confirmed using the set of genuine expressions of the new actors. In Experiment 12, TD participants showed a very similar overshoot response bias in the fake and genuine Joy-to-neutral video-clips. This result may have been due to a failure to implicitly recognize the difference between genuine and fake smiles. Therefore in Experiment 13 participants were informed through superimposed labels whether the video-clip depicted a genuine or a fake expression (fake angry facial expressions were included as well). Participants were further instructed that the fake expressions were not linked to any specific emotional state of the actor. However, the overshoot biases in the fake and genuine conditions did again not differ from each other. Possibly, the absence of a difference between the faked and genuine facial expression experiments indicates that the mechanism responsible for emotional anticipation is not sensitive to explicit knowledge, but instead proceeds in an involuntary manner on the basis of the literal/featural information contained in the immediate perceptual history.

Experiment 14 consisted of an electromyographical study (reported in the Appendix) in which activity of the Zygomaticus Major (smile muscle) and the Corrugator Supercilii (frown muscle) was recorded from participants while they observed the Anger/Joy-to-neutral video clips and judged the final experessions. The results suggested that, at least for a subset of the participants (9/15), the overshoot response bias in the Joy-to-neutral condition was associated with increased activity in the Corrugator, while the overshoot bias in the Anger-to-neutral condition was associated with increased activity in the Zygomaticus. These results are preliminary, but so far seem compatible with an embodied simulation account of emotional anticipation.

Overall, the findings presented in Chapters 2, 3 and 4 suggest that dynamic presentations of facial expressions generate in TD individuals an 'emotional anticipation', which can be seen as a low-level mindreading mechanism possibly involving a motor simulation through mirror mechanisms, that in turn shapes the perception of those expressions. In AS individuals the 'emotional anticipation' seems to be compromised and may in the current tasks have been compensated for by relying on sequential contrasts. Implications for the neural mechanisms underlying emotional anticipation, and its dysfunction in AS, are discussed within the theoretical frameworks of action understanding through (1) embodied simulation and 'direct matching' mirror mechanisms, and through (2) inferential reasoning.

Acknowledgments

I would like to thank my supervisor Dr Tjeerd Jellema for his precious support in guiding me through my PhD course and my thesis. I am grateful to members of the Psychology Department at the University of Hull, especially my colleague and friend Hollie Burnett for help in data collection, Daniel Walters for long dialogues on the research topic, Dr Johan Hulleman for statistical advice and Dr Chris Longmore for teaching some tricks of the trade...

I am grateful to Prof David Perrett in St. Andrews University for providing the morphed pictures used for the experiments presented in Chapter 2 and 3. I would also like to thank Prof Vittorio Gallese and Dr Rachel Wood in the Department of Neuroscience at the University of Parma, Italy, for their guidance during my EMG training.

Many thanks to the actors, the photographer and the artistic director, Eleonora Kojucharov, of the cultural association ‘Accademia Arvamus’ in Rome, Italy for having made possible the creation of the ‘fake’ facial expressions.

And last but not least, many thanks to my boyfriend, my family and all my friends for their unconditional understanding.

Declaration

This thesis contains original work completed by myself under the supervision of Dr Tjeerd Jellema.

The data presented in Chapter Two have been submitted and/or published in:

Palumbo, L., & Jellema, T. Beyond face value: involuntary emotional anticipation shapes the perception of dynamic facial expressions. (under review).

Jellema, T., Pecchinenda, A., Palumbo, L., & Tan, E. G. (2011). Biases in the perception and affective valence of neutral facial expressions induced by the immediate perceptual history. *Visual Cognition, 19*(5), 616-634.

Palumbo, L., Burnett, H., & Jellema, T. (2011). Emotional anticipation shapes social perception. *Perception* 40 ECVP Abstract Supplement, 214.

Jellema, T., & Palumbo, L. (2010). Facial expressions and emotional anticipation. *Perception,* 39 ECVP Abstract Supplement, 94.

And were presented at the following conferences:

Jellema, T. & Palumbo, L. (2011). Emotional anticipation in high-functioning autism. *European Society for Cognitive Psychology (ESCOP)*. San Sebastian, Spain. September 29 – October 2.

Palumbo, L. & Jellema, T. (2011) Emotional anticipation shapes social perception. *European Conference for Visual Perception (ECVP)*, Toulouse, France. August 28 -September 1.

Palumbo, L. & Jellema, T. (2011). “Beyond face value: involuntary ‘emotional anticipation’ influences social perception.” *Experimental Psychological Society (EPS*), Nottingham, UK. July 6-8. (Talk).

Palumbo, L. & Jellema, T. (2010). Facial expressions and emotional anticipation. *British Psychological Society (BPS),* Cognitive Psychology Section. Cardiff, UK.

Jellema, T. & Palumbo. (2010). Facial expressions and emotional anticipation.” *European Conference for Visual Perception (ECVP)*, Lausanne, Switzerland. August 22-26.

Jellema, T. & Palumbo, L. (2010). The Use of Social Cues for the perception of Others' actions and perspectives”. *1st joint Experimental Psychological Society (EPS) and SEPEX conference*, Granada, Spain. April 15-17.

Palumbo, L., Walters, D. Overgaard, S., & Jellema, T. (2009). Emotional anticipation”. *17th European Society for Philosophy and Psychology (ESPP)*. Central European University, Budapest. Hungary. August 27-30.

Table of Contents

List of Figures

List of tables

CHAPTER 1

Implicit, automatic and explicit inferential mechanisms in social understanding

Humans live in a complex social environment full of reciprocal interactions, which require sharing intentions, interests and beliefs, and communicating motivations and emotions. Especially non-verbal cues, such as body gestures, eye gaze direction and facial expressions are thought to be direct signal of other people's inner states.

The interpretation of others' actions in terms of an '*intentional stance*' is a pivotal component of social cognition because it enables intersubjective understanding (Dennett, 1987; Frith & Frith, 2003). Typically developed humans are equipped with a Theory of Mind (ToM), which is the ability to understand and predict a full range of mental states, such as beliefs, desires, motivations, and emotions (Baron-Cohen, 1995). Thus, in order to engage with others and respond appropriately to different social requests, people need to comprehend others' mental life. Individuals who have difficulties in understanding the experience of others, as is often proposed to be the case in autism, find it much harder to face the demands of day-to-day social life (Baron-Cohen, 1995).

The process with which humans come to understand their conspecifics and anticipate their behaviour is still much debated in distinctly different disciplines, such as psychology, social cognitive neuroscience and philosophy of mind. One of the challenging questions is whether social understanding basically derives from automatic, involuntary processing of social information, or whether more deliberate, controlled or reflective modules are responsible for the social understanding, or, whether these two components work in concert (Adolphs, 2009; Frith & Frith, 2008; Lieberman, 2007; Satpute & Lieberman, 2006).

1.1 Early development of the ability to anticipate other's actions

Several theories have been proposed to explain how people make sense of and anticipate others' behaviour (see section 1.2). These can be broadly differentiated in those that employ the embodied mind thesis for explaining action understanding and those that refer to the use of a set of theories or concepts and additional cognitive resources to make inferences about other people's mental states. These accounts base their positions on a prior differentiation between a mechanistic and mentalistic analysis of social events.

While watching a scene in real life one of the major distinctions made by the visual system is between agentive, animate stimuli, whose behaviour is self-propelled and initiated by goals and intentions, and inanimate, non-agentive stimuli, which lack inner mental processing. The motion of inanimate, non-agentive objects is processed in terms of physical laws, or input - output relationships, underlying mechanical causality, as is the case for collisions between billiard balls or moving geometrical shapes on a display (Michotte, 1963). However, a movement of an object which is self-propelled is perceived as agentive (Heider & Simmel, 1944; Premack, 1990). In other words, the observation of movements which are self-initiated and goal directed (i.e. actions) leads to the attribution of agency and intentionality (Baron-Cohen, 1995, 2005). This is evident from experiments using moving point-lights located at the joints of the actor presented on a display, which are immediately perceived as a human action, e.g. walking or running. This phenomenon has been termed 'biological motion', referring to a self-initiated, intentional movement of objects (Johansson, 1973; Kozlowski & Cutting, 1977). It was replicated with different kinds of human actions, including social and instrumental actions (Dittrich, 1993). Further, bio-motion perception seems to facilitate social judgments, such as those concerning the identity of actors (Lou & Baillargeon, 2005) or emotional expressions (Walk & Homan, 1984; Dittrich, Troscianko, Lea, & Morgan, 1996). Humans spontaneously allocate visual attention to biological stimuli, such as human bodies, compared to other stimuli (Downing, Bray, Rogers, & Childs, 2004; Thornton & Vuong, 2004). This preference seems to be present from birth (Simion, Regolin, & Bulf, 2008) and it may promote the development of higher-order social competencies.

At early stages of the development infants are argued to perceive moving objects as agentive and having goals by teleological means, independently of whether they are self initiated or not (Csibra & Gergely, 1998). For instance, one-year olds attribute goals not only to people or puppets but also to computer-animated shapes (Gergely & Csibra, 2003). These findings led some to propose that estimations of an action's goal are achieved by 'teleological reasoning'.

According to this model infants are more likely to predict the goal of an action (i.e. end-point of a moving ball) when the agent tries to achieve the goal in the most efficient way available and with the least loss of energy, also when this involves the use of impossible biological actions (Csibra, 2007). The processing of the mechanical properties of objects, according to Leslie (1994a), is the first level of the architecture for the cognition of agency. By 18 months of age, infants' interpretations about the actions of people are different from inferences made about the motions of mechanical devices (Meltzoff, 1995). Thus, this next step in the development refers to the formation of concepts about people as intentional agents (Woodward & Summerville, 2000), representing the goal of their actions as the content of their mental state.

The real social environment is characterized by complex interactions between animate agents, and inferences and predictions about others' intentions may derive from salient social or behavioural cues (Grèzes, Frith, & Passingham, 2004; Pelphrey, Singerman, Allison, & McCarthy, 2003; Saxe, Xiao, Kovacs, Perrett, & Kanwisher, 2004), especially facial gestures and eye gaze direction which specify intentional/social contingency (Pelphrey, Morris, Michelich, Allison, & McCarthy, 2005). Humans are required to identify and interpret those social signals in order to response appropriately to different social demands.

These inferences do not only concern the attribution of a goal to someone's action (first order intentionality), but also concern the other person's inner states which incorporate intentions, attitudes, beliefs, motivations and emotional states (second order intentionality).

Thus, in the social scenario understanding why people do what they do and to anticipate their actions might represent a further step in the analysis, which is not only based on the visual description of the social events themselves (Jellema & Perrett, 2007). The detection of agency, based on movement and contingency, might support higher-order, inferential processes about inner mental states (Blakemore et al., 2003).

1.1.1 The neural basis for biological motion and action anticipation

Studies on humans and non-human primates revealed the existence of a specialized brain network that processes biological motion. This system includes cortical regions in the both the dorsal and ventral visual streams. The former deals with object location and motion ('where' region), while the latter is associated with object recognition (the 'what' region) (Goodale & Milner, 1992). It has been shown that two specific regions, the Fusiform Body Area (FBA) and the Extrastriate Body

Area (EBA) in the human posterior visual cortex, are selective to the sight of body parts (Downing, Chan, Peelen, Dodds, & Kanwisher, 2006; Peelen & Downing, 2007) and are involved in the analysis of explicit low-level visual features of human bodies (Downing & Peelen, 2011).

Specifically, the FBA, located ventrally in the fusiform gyrus (Peelen & Downing, 2005), is more responsive to the whole body as compared to individual body parts, while the EBA, in the posterior inferior temporal sulcus/middle temporal gyrus (Downing, Jiang, Shuman, & Kanwisher, 2001), is more sensitive to the amount of body presented (Taylor, Wiggett, & Downling, 2007). Importantly the EBA does not respond to the correct temporal sequence of frames, thus is more likely to process the form and the structure of the body rather than biological motion (Downing, Peelen, Wiggett, & Tew, 2006). Moreover, EBA does not distinguish between actions performed by others and those of oneself (Urgesi, Candidi, Ionta, & Aglioti, 2007; Astafied, Stanley, Shulman, & Corbetta, 2004). The EBA projects to the superior temporal sulcus (STS; Giese & Poggio, 2003), which is located in the lateral posterior temporal cortex of the human brain.

The STS region is responsive to bodily postures presented in a coherent temporal order and is involved in the processing of causal contingencies (Blakemore & Decety, 2001). It is well documented that the STS plays a functional role in processing biological motion. The human STS showed activation for the perception of biological motion of various human body parts, such as hand actions (Grafton, Arbib, Fadiga, & Rizzolatti, 1996), eye gaze direction and mouth movements (Hoffman & Haxby, 2000; Puce, Allison, Bentin, Gore, & McCarthy, 1998), static faces (Allison, Puce, & McCarthy, 2000), and meaningful actions (Decety & Grezes, 1999, Decety et al., 1997; Saxe et al., 2004). Recent neuroimaging studies in humans have confirmed that the STS may play a role in representing goal-directed or intentional actions (especially in the right hemisphere; Pelphrey et al., 2005; Saxe et al., 2004). However, the STS does not seem to account for any motivational inferences, such as desires, fears and beliefs that may motivate the action. Moreover, the STS has connections with the amygdala (Aggleton, Burton, & Passingham, 1980) and the orbitofrontal cortex (Barbas, 1988). These areas are involved in the processing of social and emotional stimuli in humans and non-human primates (see Brothers, 1990; Adolphs, 1999).

Single cell studies in the Macaque monkey have found STS cell populations selectively responsive to specific bodily actions, eye gaze direction as well as head and torso orientation. These activations were reported either when those behaviours were performed in isolation or in specific combinations (Perrett, Hietanen, Oram, & Benson, 1992). Other STS cell populations were found to be involved in coding the end-point of an action sequence based on the immediately preceding movements (Jellema & Perrett, 2003).

Importantly, in order to respond some STS cells required the additional presence of a bodily or environmental a cue (i.e. preceding movements or gaze direction or reaching for a cup; Jellema, Maassen, & Perrett, 2004). Such STS cells seem to code the action when it is embedded in context (Jellema, Baker, Wicker, & Perrett, 2000; Jellema et al., 2004; Perrett, 1999). All together the above findings suggest that the function of the STS may be to provide a description of a bodily action in terms of its causes, goals and consequences (Jellema & Perrett, 2007), and how it evolves in the immediate future (Perrett, Xiao, Jellema, Barraclough, & Oram, 2006; Perrett, Xiao, Barraclough, Keysers, & Oram, 2009). Therefore, the STS may be involved in the mechanistic, contextual and probabilistic aspects of goal-directed and intentional actions (Jellema & Perrett, 2005).

1.2 Social understanding: Theory Theories and Simulation Theories

The main attempt of classical cognitivism was to establish how mental processes are generated and the causal relation between mental states and behaviour. According to the classical view of cognitivism, cognition is explained in terms of computational processes (TOTE: text, operate, text exit; Miller, Galanther, & Pribram, 1960), which generate symbolic representations in the mind (Fodor, 1981; Pylyshyn, 1987). In this scenario two different theoretical accounts tried to explain how people can understand the behaviour of others in terms of epistemic mental states. In line with the classic view of cognitivism, Theory Theory (TT) argues that common-sense terms for mental states, such as intentions, beliefs, desires, motivations and emotions, are part of a theoretical framework. In other words the way in which people grasp others' mental states requires the use of a theory. Then, having a theory of mind or the ability to read someone else's mind would invoke laws and inferences (Gopnik & Wellman, 1992; Leslie, 1994a).

Simulation theories were inspired by experiments on visual imagery. For example, Shepard & Metzler (1971) found that rotating an object mentally takes roughly the same time as for the object to rotate in the reality. This indicated that there is much in common between pretended and real-generated states. During simulation the system is taken off-line so that the outcome is not actual behaviour but predictions and anticipations of the behaviour of others. Thus, according to Simulation Theory (ST) the basic idea is that mental states may be attributed to others by creating pretend states in oneself, or putting oneself into the other's shoes (Goldman, 1989; Gordon, 1995; Heal, 1986). Advocates of ST claim that mental states concepts derive from experiencing and simulating such mental states when explaining and predicting the same mental states in other people. Therefore, in contrast to TT accounts, simulation would immediately provide people with the meaning of others' behaviour, without the need of inferential reasoning or the application of a theory.

Both TT and ST share the assumption that mental states are unobservable and that the only information humans rely on to understand others' minds is accessible by the observation of their behaviour (Herschbach, 2008). What is debated by the two theoretical models is whether such access is theory-driven, i.e. achieved by applying theoretical knowledge via explicit or implicit inferences (Leslie, 1994a; Fodor, 1968; Stich & Nichols, 1992), or whether it is process-driven (Goldman, 1989). In the next paragraphs contributions and limitations of the two theoretical accounts (TT and ST) will be discussed in relation to the current debate on ToM.

1.2.1 Theory Theory (TT)

Advocates of TT argue that mindreading is a theorizing activity. People understand other's mental states by making attributions and utilizing folk psychological theories (Churchland, 1991), which are informative about the most probable causes of other's people behaviour. Thus, understanding others would require adopting a theoretical stance and psychological laws, which define the link between inner mental states and behaviour.

Alison Gopnik described the acquisition of these high-order cognitive activities in developing children using the example of a scientist acquiring knowledge, known as the 'child-scientist-theory'. Gopnik argues that theories of mind are not innate but that the child acquires folk psychology by using the same causal reasoning mechanisms as those used by a scientist (Gopnik & Wellman 1994; Gopnik &

Meltzoff 1997). Thus, third-person attribution is purely inferential and the acquisition of mental-state concepts derives from causal relations between behaviour, environment and inner states, which relations are theoretically specified and constantly updated.

In contrast, the innatist view suggests that humans are genetically predisposed to develop concepts such as belief and desire. This innatist view is related to the modularity theory proposed by Fodor (1978, 1983). The idea is that people hold a system of inferences and ‘representational relations’ (i.e. pretends, believes, desires that are involved in the causation of the agent’s behaviour), which together constitute a tacit, implicit and intuitive ToM (Leslie & Roth, 1993).

According to this account the ToM mechanism relies on these innate concepts in combination with the predisposition to direct attentional resources to other people’s mental states (Scholl & Leslie, 1999).

1.2.1.1 ToM tasks

The traditional view of ToM as a domain-specific theory suggests that in order to explain and predict the behaviour of others one needs to posses mental states concepts. According to TT advocates the course of development is seen as a maturation of child’s ToM concepts which contribute to the formation of a folk-psychology theory in adults. Empirical tests for ToM first focused on assessing whether and at which stage in the development children start to attribute false beliefs, i.e. the ability to recognize that others can have beliefs which are different from their own beliefs. In the ‘Sally-Anne’ task, developed by Wimmer & Perner (1983), children are told, or shown, a story about Sally and Anne. While Anne is watching, Sally puts a marble in her basket and leaves the room. While she is away, Anne takes the marble from the basket and puts it in a box. When Sally returns, the child is asked where Sally will look for the marble. Thus, the participant is asked to report the content of the actor's belief. The right answer requires attribution of false belief to Sally which is different from one's own mental representation of the situation. It was found that most children before age of four are unable to respond correctly.

This finding was interpreted as a lack of a false belief concept (Gopnik & Wellman, 1992; Wellman, Cross, & Watson, 2001). In line with the child-scientist approach, correct belief attributions are instantiated by the development of theoretical reasoning, which promote the understanding of mental concepts.

In a modification of the Sally-Anne task (Baron-Cohen, Leslie, & Frith, 1985), 18 months-old infants looked significantly longer when an agent's behavior was incongruent than when it was congruent with the agent's false belief. This result contradicts the view that theory of mind requires belief concepts at around 4 years of age (e.g., Gopnik & Wellman, 1992; Ruffman & Perner, 2005; Saxe, Carey, & Kanwisher, 2004).

In the last three decades, it was long debated whether failures in the false-belief tasks with young children reveals conceptual deficits or information processing biases (Mitchell & Lacohee, 1991; Lewis, Freeman, Hagestadt, & Douglas, 1994). It was also suggested that the saliency of the information provided, i.e. the 'true' location of the toy at the beginning of the task, could have guided their response (Zaitchik, 1991), or that childrens' failure of the false-belief tasks derives from a difficulty to inhibit their previous knowledge and update it accordingly to the other's person desire-change (Moore et al., 1995). Other issues concerned the involvement of verbal reasoning in the standard tasks (Siegal & Beattie, 1991) or additional abilities other than understanding mental states (Bloom & German, 2000). It was suggested, indeed, that the attribution of beliefs may follow an intuitive, automatic form of reasoning that may be independent of language (Friedman & Leslie, 2004).

Notably, when advanced versions based on non-verbal tasks were employed, the development of false-belief attributions was found before four years of life. Onishi & Baillargeon (2005) combined a violation-of-expectation method with a preferential looking paradigm (Woodward, 1998). They showed an adult putting an object in one of two boxes to 15-months old infants. An occluder was inserted hiding the agent's view of the two boxis. The object was then moved from one box to the other one. Once the occluder was removed, the agent could either reach the original box, accordingly to the agent's false belief or the new location (incongruent choice to the agent's false belief). The authors found that 15 months-old infants looked significantly longer in the latter condition, thus suggesting that infants before four years old already posses false belief concepts. These findings were further corroborated by Southgate, Senju & Csibra (2007) who found that two years old children could already anticipate someone else's actions when the action's prediction was only possible by attributing a false belief to the actor.

According to those findings, it seems that children under four years old hold ToM concepts, and that the lack of correct answers on standard false belief task was due to the use of direct, explicit measures. However, having mental states concepts does not necessarily incorporate the ability to use them efficiently in mindreading (Apperly, Samson, & Humpreys, 2009; Samson & Apperly, 2010). This can explain why also adults make errors in ToM tasks, even though they already hold those concepts (Keysar, Lin & Barr, 2003).

Thus, it is suggested that additional cognitive resources implicated in the inhibition of one's own egocentric perspective and in the selection, monitoring and integration of salient cues to reason about the mental state of that particular person, are necessary to solve ToM tasks (Apperly et al., 2009; Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006; Samson & Apperly, 2010). Thus, the task may appear more demanding when subjects, children or adults, are required to overcome their own egocentric view and inhibit the self perspective to take the perspective of another. Interestingly, there is evidence that adults can implicitly process someone else's visual experience even when they themselves hold a different point of view. However, this would only occur when there is no need to refrain from one's own point of view (Samson, Apperly, Braithwaite, Andrews, & Scott, 2010). In the latter case the other person is not seen as a different entity than one self and therefore there is no need to inhibit one's own perspective (Samson & Apperly, 2010). This may also explain why infants before four years of age were able to solve the 'implicit' versions of the false belief tasks (Onishi & Baillargeon, 2005; Southgate, Senju & Csibra 2007).

The question is whether in every social interaction people would need to apply such cognitive computations to understand how the other person is feeling or which intention underlies her/his behaviour. If people would always need to recruit such effortful cognitive resource when engaging with others, then being successful in social interactions would be hard.

The development of the ability to understand mental states other than belief and desires was also investigated by considering imitative skills and joint attention. For instance, Meltzoff (1995) tested the ability to replicate and complete unfinished goal-directed acts. He found that 18-month-olds can grasp the meaning of others' goals even when the attempts to achieve those goals fail. The ability to imitate the intended action, as well as completed actions, at this age was confirmed by Carpenter, Akhtar, & Tomasello (1998). Joint attention is considered a precursor to the development of ToM (Baron-Cohen, 1989; 1991a). By 14 months of age children engage in joint attention, they share the attention toward an object with the adult's attention, by following her gaze or pointing gestures (Tomasello & Haberl, 2003). At earlier stages, between 4 and 9 months, infants start to develop the ability to pick up social cues, including eye gaze and facial expressions, and discriminate between those signals which are directed to the self and the other (i.e. dyadic context) and those which are directed to the environment (triadic context) (Striano & Stahl, 2005). For instance, at 7 months of age infants show sensitivity to referential fearful facial expressions (with eye gaze directed to an external object) suggesting an early ability to use relevant social information to process potential dangerous objects in the environment (Hoehl, Palumbo, Heinisch, & Striano, 2008).

Thus, the ability not only to detect but also to use social signals properly while engaging in social interactions starts to develop very early in life, despite that at that stage high cognitive functions are still poorly developed.

In conclusion, on one hand new directions tend to emphasize the role played by cognitive functions and high level reasoning processes involved in the use of such concepts, especially in those tasks where taking the perspective of the other person is an effortful activity. On the other hand, it is possible that social understanding may also be supported by implicit processes, which can make use of immediate social information without requiring such difficult cognitive computations. Whether and when in real life people adopt one or the other way this possibly depends on the complexity of the social scenario.

1.2.1.2 The neural correlates of ToM: the mentalizing network

It was suggested that a specific neural network is involved during tasks assessing the comprehension of beliefs and desires (Frith, 2001; Frith & Frith, 2003; Saxe & Kanwisher, 2003). This network, which includes the medial prefrontal cortex, the temporal parietal junction (TPJ) and the precuneus, is commonly termed the 'mentalizing network'. The TPJ is located around the supramarginal gyrus; the medial prefrontal cortex is considered the medial area of the prefrontal cortex and the precuneus is located in the posterior medial brain. Typically, this network is involved in tasks assessing explicit inferential components of social understanding. For instance, the mentalizing network is involved when participants are asked to assess other people's intentions from stories (Jenkins & Mitchell, 2010), using pictures of human actions (de Lange, Spronk, Willems, Toni, & Bekkering, 2008) or displays of moving geometrical shapes (Castelli, Happe, Frith, & Frith, 2000). Recently the same areas where found active when participants observed irrational actions as compared to rational actions (Brass, Schmitt, Spengler, & Gergely, 2007). However, another study did not find the same engagement of TPJ during irrational actions (Jastorff, Clavagnier, Gergely, & Orban, 2010). In a recent meta-analysis study the mentalizing areas were found active when inferences about goals, beliefs or moral issues are presented in an abstract way and without involving biological motion of body parts (Van Overwalle & Baetens, 2009). According to this meta-analysis, the TPJ results involved in understanding the intent of a social agent when goal-directed social behaviour do not include body-part information. The medial prefrontal cortex seems to be recruited in deliberative reasoning or trait inferences (Van Overwalle & Baetens, 2009).

1.2.2 Simulation Theory (ST)

In contrast to TT accounts, defenders of the simulation approach claim that pre-linguistic representations of mental states are not generated by inferential processes or high level theorizing, but rather by creating simulated ‘pretend states’ in the self, *as if* the observer would be standing in the agent’s shoes (Goldman, 2006; Goldman & Sripada, 2005).

In *Simulating minds* Goldman (2006) provides a distinction between low-level mindreading and high-level mindreading. The former is supported by an automatic, unconscious simulation (or ‘unmediated resonance’, Goldman & Sripada, 2005), which is free from any task-specific knowledge (low-level simulation). The second form of mindreading is supported by high-level simulation which is linked to the traditional version of pretence-driven or imagination-driven simulation and it is driven by task-specific information (Goldman, 2009). Goldman suggests that low level mind-reading supports motor intention attribution and face-based emotion attribution, while the high-level mindreading may better explain decision, desire and belief attributions.

Although the notion of simulation is considered misleading as it may assume different meanings (Gallagher, 2007; Stich & Nichols, 1992), here simulation of the other person’s states of mind is meant to be a functional process which contains representations of the possible mental/emotional states of target objects (Gallese & Lakoff, 2005; Gallese & Sinigaglia, 2011; Goldman, 2006).

1.2.2.1 Low level mindreading, embodied simulation and the ‘mirroring’ mechanism

Following the work: ‘*Ecologic Approach to Visual Perception’* by Gibson (1979), a new view of cognitive science, called Embodied cognition came about which emphasized the importance of body-environment interaction and its role in the production of internal mental states (Lakoff & Johnson, 1999). This approach involved several contributions from different disciplines ranging from philosophy of mind and cognitive neuroscience to artificial intelligence and robotics. Embodied cognition emphasizes the way cognition is shaped by the body and its sensori-motor interaction with the surrounding social and material world, making clear that mental operations involved in cognitive activities are not computed only in the mind (Clark, 1997; Clark & Chalmers, 1998; Thelen & Smith, 1994). In philosophy of mind the phenomenological tradition already provided an account for embodiment and its relation to cognition (Gallagher, 2005; Lennon, 1990; Lennon & Gilbert, 2005). For instance, Heidegger in his '*being-in-the-world'* put light on the fact that humans are engaged in dealing with things in the world which then allows a theoretical constructions of the reality. Similarly, Merleau-Ponty (1962) observed: "*Representations are therefore 'sublimations' of bodily experience, possessed of content already, and not given content or form by an autonomous mind; and the employment of such representations "is controlled by the acting body itself, by an 'I can' and not an 'I think that'* ".

In line with theories of Embodied cognition on the embodiment of the mind, a new alternative model of simulation, called *embodied simulation,* has been proposed to explain intersubjective understanding (Gallese, 2001, 2007). According to Gallese (2007), humans are equipped of a simple, direct and automatic mechanism enabling intersubjective understanding in a very immediate way and without the use of complex reasoning or inferential analogy. Embodied simulation, which is the substrate of low-level mindreading as discussed by Goldman (2006), refers to an activation of neural mechanisms in response to someone' else behaviour, which are the same as those used in the production of one's own behaviour. Thus, understanding others' mental states can be explained in terms of sensori-motor processes through the re-activation of neural circuitry active in bodily perception, action, and emotion. The neural mechanism enabling such direct understanding is often referred to as the Mirror Neuron System (MNS; Rizzolatti & Craighero, 2004).

Mirror neurons in the monkey's brain

Mirror neurons (MNs) are a particular class of motor neurons, which were first discovered in the ventral premotor cortex (area F5) of the macaque's brain using single-cells recording (Di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996).

These neurons, which constitute around 20% of cells in F5, are called 'mirror neurons' because they discharge both when the monkey performs an action (such as grasping) and when the monkey observes a similar action done by another monkey or the experimenter. Importantly, when the motor act was executed in the absence of a target object (e.g. food) these mirror neurons were silent. This led to the idea that mirror neurons could underpin action understanding (Rizzolatti, Fogassi, & Gallese, 2001). The mirror neuron system (MNS) in the monkey's brain also includes the inferior parietal lobule (IPL; Fogassi et al., 2005; Gallese, Fadiga, Fogassi, & Rizzolatti, 2002). The latter receives input from the STS (Rozzi et al., 2006), which responds only to the observation of biological movements, and sends an output to F5 (see Keysers & Perrett, 2004 for a review). Mirror neurons in F5 were classified in two main categories: 'strictly congruent' and 'broadly congruent'. In the first category there are mirror neurons (one third of all mirror neurons) that in order to be triggered require the observation of the specific action they would execute (e.g. grasping an object with a precision grip instead of a power grip). The remaining two third of all mirror neurons are 'broadly congruent', which can be triggered by dissimilar actions (grasping with the hand or grasping with the mouth) which have the same goal. Thus, it was suggested that mirror neurons allow the observer to understand *what* has been witnessed, namely the overall goal of others' actions (Gallese et al., 1996; Rizzolatti & Craighero, 2004).

Further, mirror neurons do not fire during observation of a pantomimed action (without the target object), but they do fire if the monkey sees the starting point of an action (hand-reaching for a target object), which is subsequently occluded by a screen (Umiltà et al., 2001). Importantly, the same study revealed that if the monkey knew that there was not an object behind the screen then the discharge of the mirror neurons was absent. This finding suggested that mirror neurons activity reflects a representation of the agent's goal. Importantly, the described mirror neurons properties were not only found with transitive actions but also with intransitive, communicative mouth actions, e.g. lip-smacking (Ferrari, Gallese, Rizzolatti, & Fogassi, 2003). Thus, mirror neurons seem to underpin also social facial communication in monkeys. Moreover, it has been shown that 'audio-visual F5 MNs' discharge not only during execution and observation of a (noisy) act (e.g. breaking a peanut), but also by simply hearing its sound (Kohler et al., 2002). In another study monkeys were trained to use two types of pliers requiring either opening or closure of the hand in order to grasp food (Umiltà et al., 2008). Notably, F5 MNs in these trained monkeys responded to the observation of grasping performed with both types of pliers. This finding suggests that motor knowledge can modify MNs motor proprieties and can also modulate the way MNs generalize their visual response. Therefore, it was suggested that MNs may also code action's intention and not only its goal (the end-state of the action). In a different study the monkey was presented with two conditions in which grasping a piece of food or an object was achieved either to eat or to place (Fogassi et al., 2005). In the first part of the task grasping the target was identical in the two conditions. The results showed that almost all of the parietal and premotor MNs were sensitive to the action intention, i.e. to either eat or to place. The data led the authors to propose that motor acts are organized in *chains* (grasp-for-eating and grasp-for-placing). Importantly, these chains were activated *before* the observation of the forthcoming motor act of the chain, thus providing the observer with a map of the whole future action before it was executed. It has been suggested that these MNs could provide the neural basis for action prediction and intention understanding.

Recently, in a behavioural study a looking paradigm previously applied to human babies (Gergely, Nadasdy, Csibra, & Biro, 1995) was adapted to test the ability of macaque monkeys to evaluate and predict goal-directed actions of others (Rochat, Serra, Fadiga, & Gallese, 2008). The authors found that non-human primates, similarly to 9 to 12-month-old humans, were able to detect the goal of an action and build expectations about the most likely action the agent will execute in accordance with the physical characteristic of the context. It should be noted though that this was evident only when those actions were part of the monkey's own motor repertoire, thus suggesting that previous experience plays a crucial role.

On the basis of these results the *mirroring* mechanism is interpreted as enabling the transformation of sensorial information into potential motor acts. The derived sensorimotor information produces multimodal representations equipped with the meaning of the observed actions. Thus, according to the group of researchers in Parma, by automatically matching the agent's observed movements onto her own motor repertoire (*motor resonance*, Rizzolatti, et al., 2001) without executing them, the activity of mirror neurons in the observer's brain simulates the agent's movements and thereby contributes to the understanding of the perceived action (Gallese et al., 1996; Rizzolatti et al., 1996). The direct matching approach is therefore *retrodictive* as the underlying goal of the actor is inferred by a bottom-up mechanism activated in the motor system.

Although the mirror mechanism was discovered in the monkey's brain, subsequent studies verified its presence in the human's brain too.

Mirror neurons in the human's brain

Comparative and neuroimaging studies showed that the MNS in humans includes the premotor cortex, the lower part of the prefrontal gyrus and the posterior part of the inferior frontal gyrus (Rizzolatti & Craighero, 2004).

Although almost all evidence for the presence of mirror neurons in the human brain is derived from neuroimaging studies, the presence of the mirror mechanism in the human brain was recently verified in a single neuron recording study conducted by Mukamel, Ekstrom, Kaplan, Iacoboni, & Fried (2010). Interestingly, in this study MNs were found outside the traditional circuit, thus suggesting that the MNS in humans is more broadly extended compared to the MNS in the monkeys. Importantly, in this study 'super mirror neurons' were identified and described as the cells that inhibit the execution of the observed actions, thus enabling off-line simulation. Furthermore, this subset of mirror neurons may help the observer to distinguish the actions of other people from her own actions.

The MNS is closely connected to the posterior STS which sends the visual information forward to the parietal MNS and further to the frontal MNS where such information is translated in potential motor acts. The core system is also defined as Action-Observation network (AON; Cross, Kraemer, Hamilton, Kelley, & Grafton, 2009; Kilner, 2009) referring to the action-observation matching accomplished by the MNS. The first study assessing the presence of the mirror mechanism in the human brain was conducted by using transcranic magnetic stimulation (TMS) (Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995). The left motor cortex (area 44 or Broca area, the homologue of F5 in the monkey's brain) was stimulated using TMS in order to produce motor evoked potentials (MEPs), which were recorded from the right hand muscles. Crucially, Fadiga and colleagues found a significant increase of the MEPs during observation of transitive actions that involved the agent's right hand, but not by the observation of other actions not involving the agent's right hand. It has also been shown that the mirror neuron system in humans is directly involved in imitative behaviour (Iacoboni et al., 1999; Iacoboni et al., 2005) and in the intention to imitate (Frey & Gerry, 2006).

Additional evidence for the mirroring mechanism derives from studies employing electroencephalography (EEG) and magnetoencephalography (MEG). It had been known for some time that EEG rhythm recorded from central derivations (mu rhythm) is disrupted/reduced during active movements (Berger, 1929), but it turned out that the mu rhythm is also disrupted during passive observation of actions performed by someone else (Altshuler, Vankov, Wang, Ramachandran, & Pineda, 1997; Altshuler et al., 2000). Moreover, desynchronization of the rhythmic oscillation was found during action execution and observation in the central sulcus (Hari & Salmelin, 1997; Hari et al., 1998). The intensity of activation of the MNS is correlated with the familiarity of the observed action in adults (Buccino et al., 2004; Calvo-Merino, Glaser, Grezes, Passingham, & Haggard, 2005; Cross, Hamilton, & Grafton, 2006) and infants (van Elk, van Schie, Hunnius, Vesper, & Bekkering, 2008). This finding seems to be in line with behavioral tests in monkeys (Rochat et al., 2008). Thus it seems that if the mirror system has a role in action understanding, then this is more likely the case with familiar actions which belongs to the motor repertoire of the observer.

Interestingly, in an fMRI study the hypothesis that the mirror system may serve intention understanding (Fogassi et al., 2005) was corroborated in humans (Iacoboni et al., 2005). Participants were presented with three kinds of stimuli: grasping hand actions without a context; context only (a scene containing only objects); and grasping hand actions embedded in contexts. In the context condition it was possible to discriminate the intention behind the grasping action (either drinking or cleaning up). The results revealed a significant increase of the signal in the posterior IFG and the ventral PMC when actions were embedded in contexts, compared with the other two conditions. This suggested that the mirroring mechanism may enable not only action recognition (‘what’, goal understanding), but also the intention which generated the action (‘why’, intention understanding).

Importantly, the instructions aimed at determining the intention of the observed actions did not modulate the effect observed in the premotor mirror areas. This finding, according to Gallese (2007), further suggested that at least for the simple actions employed in this study, the ascription of intentions proceeded by default, underpinned by a mandatory activation of embodied simulation.

Altogether these studies echoed the idea that mirror neurons could serve action understanding (Gallese & Goldman, 1998) or ‘teleological sensitivity’ (Decety & Grezes, 2006), imitation (Jeannerod 1994; Iacoboni, 2009) and intention understanding (Iacoboni et al., 2005).

In conclusion, it is suggested that a mirroring mechanism promoting a direct grasping of the other’s sensori-motor experience is also present in humans and it is possibly more flexible and widespread in the human brain as compared to the monkey’s brain. It is further considered that this direct action-observation module, underpins, through embodied simulation, the understanding of intentions for simple actions, especially when the observed motor acts belong to the motor experience of the subject (Gallese, 2007).

1.2.2.2 Embodied simulation and emotional stimuli

The reflexive mirroring mechanism engaged in action understanding has also been suggested to underpin the understanding of others’ emotional states (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003; Keysers et al., 2004; Leslie, Johnson-Frey, &

Grafton, 2004). Evidence for this notion comes from studies in which other people's facial expressions were observed, which showed that motor programs and visual codes for facial expressions are closely linked (Botvinick et al., 2005; Carr et al., 2003; Wicker et al., 2003). In a fMRI study Carr et al. (2003) found that premotor and parietal cortices were both involved in both facial expression observation and execution, giving support to the *direct matching* hypothesis in understanding facial expressions. Further, Wicker et al. (2003) found that the insula was activated both when subjects observed and when they experienced the emotion 'disgust'. Neurophysiological studies, using the facial Electromyographical technique (EMG), indicate that the perception of facial expressions elicits facial muscular activity congruent to the presented facial expressions (Dimberg, 1982; Dimberg & Thunberg, 1998; Larsen, Norris, & Cacioppo, 2003; Sato, Fujimura, & Suzuki, 2008). Interestingly, these reactions can even be evoked unconsciously (Dimberg, Thunberg, & Elmehed, 2000) and they may reflect the neural mechanism allowing people to understand the emotion conveyed by the face. These results found a convincing explanation in the *embodied simulation* model proposed by Gallese (2007). Thus, the observation of the target's facial expression 'directly' triggers (sub-threshold) activation of the neural substrate associated with the experience of the emotion in question. Therefore, on the basis of the feeling or experience generated by the simulated expressions, the observer's brain 'knows' (reflexively) the meaning of those expressions (Gallese, 2006a).

In summary, the above studies seem to confirm the existence of a specific low-level mindreading mechanism for face-based emotion instantiated by embodied simulation (Goldman, 2006; Goldman & Sripada, 2005; Gallese, 2007).

1.2.3 Alternative models to the 'direct matching' hypothesis for action understanding

The discovery of the mirror neurons and their contributions to low-level mind-reading and social cognition are highly debated and several aspects of the direct matching account underwent intense scrutiny.

One of the most criticized points is that the original direct matching hypothesis is not sufficient to explain the fact that two identical actions may represent two different meanings (like grasping a glass to drink or grasping a glass to clean it). Many have argued that context information and prior knowledge and some kind of psychological theory about others are necessary ingredients too. The distinctive characteristic of the mirroring mechanism is that the simulation of the motor acts is initiated by a direct matching between the acts observed in the reality and the internal simulations.

However, people can understand other's actions without the need to imitate or simulate someone's motor act exactly as it is performed, or by the same means. In a fMRI experiment it was shown that 'mentalizing' areas were active when participants were required to understand the intention of unusual actions (Brass et al., 2007). This finding may suggest that the embodied simulation account is basically confined to familiar actions or actions executed in simple, stereotypic contexts.

Several doubts were raised up about the suitability of the mirroring mechanism for explaining social understanding (see Csibra, 2007; Csibra & Gergely, 2007; Jacob, 2008; Jacob & Jeannerod 2005). It was claimed that the study conducted by Iacoboni et al. (2005) does not exhaustively exclude that mirroring mechanisms could be instantiated (top-down) by higher cognitive processing (Jacob, 2008). In contrast to the bottom-up, retrodictive explanation of action understanding by embodied simulation or direct matching, different models were suggested. For instance, Csibra (2007) proposed that whenever an intention is attributed to an observed action, the meaning of this action first needs to be analyzed at a higher cognitive level (possibly the ‘mentalizing’ system). Subsequently, the information is sent (top-down) to the motor system (how that is done remains unclear), which accomplishes an action reconstruction via a predictive simulation. Thus, according to Csibra, action understanding is the input (and not the output) of action mirroring, which in turn generates predictive action monitoring as output. This model was inspired by studies in infancy research which do not support a ‘direct matching’ hypothesis for understanding others’ intentions (Csibra, Biro, Koos, & Gergely, 2003; Lou and Baillargeon, 2005; Onishi, Baillargeon, & Leslie, 2007). However, there is still a substantial lack of studies testing the mirroring mechanism at early stages of development and additional investigations are required.

Alternatively, Kilner, Friston, & Frith (2007) proposed the ‘predicting coding’ model which combines a Bayesian approach to theories of action hierarchy, which involve different levels of actions (from kinematic to intentions and context). According to this model prior knowledge of intentions allows to detect likely action goals via motor resonance. Thus, given the information about what the action is for, what the MNS does is to predict whether the observed actions in the reality may correspond to the predicted ones. If they do not match, then a ‘prediction error’ is projected up to the action hierarchy, thus enabling the inference of the most likely cause of the action.

In summary, what is acknowledged in the direct-matching ‘mirroring’ hypothesis is its automatic predictive property. What is denied is the possibility that understanding of intentions may derive from predicting the forthcoming new goal (retrodictive explanation). What it is suggested is that the automatic predictive property of action mirroring serves action monitoring (Csibra, 2007) or action inference via error reduction (Kilner, 2009). However, what is not clarified by these models is where and how the prior knowledge about action' intention is computed and how this information (from wherever it originates) is then able to exactly find those areas of the mirror mechanism that correspond to the observed action. Moreover, when adopting Csibra's model, the question why the brain should be equipped with such a direct-matching mechanism remains unresolved.

At present it is unknown to what extent the mentalizing network and the mirroring areas depend on each other. Some authors suggested that the MNS might inform and support the mentalizing system (Blakemore, Winston, & Frith, 2004; Decety & Chaminade, 2003; Keysers & Gazzola, 2007; Uddin, Iacoboni, Lange, & Keenan, 2007), while other authors suggest the two system operate independently (Jacob & Jeannerod, 2005; Saxe, 2005; Saxe & Wexler, 2005).

Like actions, emotions are biologically relevant stimuli which involve visceral and somatosensory reactions. Importantly, while intention, i.e. a conscious purpose of the agent, underlies an action, emotional expressions are generally displayed spontaneously and involuntarily, though they can also be voluntarily posed. In either case, they have a relevant communicative function.

1.3 Why the face is so special

A lot of research focuses on socially relevant stimuli, such as faces, and exhaustively reports that faces are processed 'specially' compared to other visual stimulus categories (Kanwisher, 2000; Spiridon & Kanwisher, 2002). Faces are attended preferentially and recognized very early in life (Bushnell, Sai, & Mullin, 1989; Morton & Johnson, 1991; Slater & Quinn, 2001; Valenza, Simion, Macchi, &

Umilta, 1996). This is because face perception involves the recognition of individuals and their mental dispositions, thus providing important information enabling social interactions.

Bruce and Young (1986) proposed a cognitive model of face recognition where two different processes serve identity and emotion recognition. According to this model, the first step of the analysis is the structural encoding, which involves the perceptual processing of the facial features and their configural relations (Calder, Young, Keane, & Dean, 2000). At a second level specialized functional models enable the recognition of specific types of information from the face such as the person identity and the expressions. Facial expressions are considered as view-centred descriptions and are analysed separately by the cognitive system for their emotional meaning.

Studies with typical subjects showed that people can label expressions of unfamiliar faces as fast as those of familiar faces (Bruce, 1986; Young, Ellis, Flude, McWeeny, & Hay, 1986). This confirmed that two independent systems underpin identity and expression recognition.

The same conclusion followed from neuropsychological studies. There are brain-injured patients who can understand emotional facial expressions, despite being unable to recognize familiar faces (Bruyer et al., 1983; Tranel, Damasio, & Damasio, 1988). The opposite dissociation was also reported, with impaired comprehension of facial expressions and an intact recognition of the identities (Adolphs, Tranel, Damasio, & Damasio, 1994; Calder et al., 1996; Nijboer & Jellema, 2012; Young et al., 1995).

Relying on the Bruce and Young (1986) model, Haxby, Hoffman, & Gobbini (2000, 2002) proposed a ‘distributed human neural system’ in which the perceptual information of facial features proceeds from the inferior occipital gyrus to the lateral fusiform gyrus (identity information) and to the STS. In the latter area, dedicated face-regions process changeable aspects of the face such as eye gaze, lip movements and facial expressions; all crucial elements for non-verbal communication. Importantly, the model emphasized the bidirectionality of the connections between its components. The core system is extended to other areas such as the amygdala, the insula and the limbic system. Adolphs (2002) incorporated also the orbitofrontal cortices which are involved in the categorization of facial expressions. Further, these cortices may retrieve associated knowledge concerning the emotion, and they may be involved, via connections to motor structures, in producing emotional responses.

Eimer & Holmes (2007), reviewing recent event-related brain potential (ERPs) studies, concluded that the analysis of emotional facial expressions relies on a complex neural network that includes a rapid, obligatory, and pre-attentive classification of emotional content as well as a subsequent deeper analysis of emotional expressions in neocortical emotion regions such as somatosensory cortex, anterior cingulate, and medial prefrontal cortex.

1.3.1 Models of emotional facial expressions

Facial expressions are social signals which can convey emotional dispositions and serve interpersonal communication (Darwin, 1872/1998; Fridlund, 1994; Ekman & Oster, 1979; Russell & Fernandez-Dols, 1997; Schmidt & Cohn, 2001). Generally those two aspects go hand in hand as at the moment an individual is expressing an emotion, even when spontaneously without intentionality, she/he also communicates the associated affective inner state. Therefore, genuine emotional facial expressions can also be seen as (spontaneous) actions having an onset, a peak and an offset. However, the social value of emotional facial expressions is more evident for the so-called social, moral, emotions such as embarrassment, guilt or pride. In this scenario, also the posed, voluntary, situational expressions, which do not have a direct link to the agent's emotional state but are intentional and culturally conditioned, play an adaptive role in modulating social interactions and communication (see Ekman, 2004). With respect to felt, emotional facial expressions, one of the most debated issues is whether emotions refer to discrete affective states or whether they are mapped onto a continuum. Emotion displays to which subjects assign labels and undergo categorical perception are considered discrete states (Ekman, 1992). However, the boundaries between emotion categories may be somewhat foggy at the level of recognition (Russell & Bullock, 1986).

There are basically two views with respect to the universality of emotional face recognition. The first states that basic facial expressions are universally recognized (Ekman, 1994; Ekman & Friesen 1971) and fall in discrete categories, with a differentiation between basic and complex emotional displays (Ekman,1992).

Specifically, it was proposed that there are seven discrete emotional expressions (neutral, happiness, sadness, surprise, fear, anger and disgust), which are biologically determined and are recognized by all individuals regardless of ethnic or cultural differences (Darwin, 1872/1998; Ekman & Friesen, 1971; Izard, 1994). Complex emotions can be considered subcategories or combinations of basic facial signals.

The second view questions about the universality of emotional facial expressions, and suggest that emotions fit into two dimensions, arousal and valence, each with a similar structure (Russell, 1980, Russel & Fehr, 1987).

According to the latter view, the categorization of the expression is not universal but depends to some extent on the contextual relation to the other expressions (Russell & Fehr, 1987).

Despite those differences, both the discrete-category view (Ekman, 1992) and the dimensional view (Russell, 1980) share the notion that the affective information are directly “read out” from the face. Specifically, facial expressions are processed by a perceptual system that analyzes both configural and facial features information to identify the emotion (Calder et al., 2000). Consequently, they are immediate signifiers of affective dispositions of other people (Adolphs, 2002, 2003; Ekman, 1992, 1993; Russell & Fernandez-Dols, 1997).

1.3.2 The influence of bottom-up visual processes in the evaluation of facial expressions

The perception of an emotional facial expression typically focuses on the facial features and their configural relations (Calder et al., 2000). The visual analysis of such physical proprieties guides (bottom-up) the perceptual evaluations of the face stimuli, corresponding to the first-step analysis in the model proposed by Bruce and Young (1986). Perceptual judgments of facial expressions can be influenced by bottom-up visual processes, as is the case for many other classes of objects. A large body of studies reported that emotional facial judgments can be influenced (bottom-up) by a range of different visual processes/mechanisms, such as contrast or context effects, adaptation and extrapolation or Representational momentum (RM).

In contrast/context effects, the presentation of a stimulus influences the perception of a subsequent stimulus (e.g. Russell & Fehr, 1987; Suzuki & Cavanagh, 1998, Tanaka-Matsumi, Attivissmo, Nelson, & D'Urso, 1995; Thayer, 1980;). Thayer (1980) showed that sad and happy facial expressions were rated as more intense when they were preceded by contrasting, as opposed to similar, facial expressions. This effect was previously observed by Russell & Fehr (1987). These authors presented two facial expressions of different identities side by side and found that the judgment of the first face (anchor) affected the judgment of the second face (target) in a repulsive manner.

This effect was particular evident when the target was a neutral facial expression. Ekman and O'Sullivan (1988) argued that although the perception of emotion in facial expressions should be accurate and absolute, neutral or ambiguous expressions may be subject to contrast/context effects. Tanaka-Matsumi et al. (1995) found that a happy anchor face would make a neutral face (the target) less pleasant and less arousing as compared to a sad anchor face. Interestingly, the fact that two different identities were depicting the emotions did not modulate the contrast/context effect.

Visual adaptation to a (distorted) stimulus produces a so called 'aftereffect', i.e. a biased perception of the stimulus towards the opposite of the adapting stimulus. Adaptation was found for various visual features such as size, orientation, curvature, motion, spatial frequency and natural images. In the literature, even visual adaptation for high-level stimuli such as faces is well documented (Leopold, O'Toole, Vetter, and Blanz, 2001; Rhodes, Jeffrey, Watson, Clifford, & Nakayama, 2003). Face after-effects induced by adaptation were found for at least five different facial traits: identity (Leopold et al., 2001), gender (Webster, Kaping, Mizokami, & Duhamel, 2004), attractiveness (Rhodes et al., 2003), ethnicity (Webster & MacLin,

1999) and facial expression (Fox & Barton, 2007; Hsu & Young, 2004; Rutherford, Chattha & Krysko, 2008; Webster et al., 2004). With respect to facial expressions, it was reported that prior presentation of happy facial expressions for 180 s (with 5 s top-ups) made subsequently presented neutral expressions look slightly angry, reflecting a shift in the boundary between the happy and angry categories in the direction of the happy expression (Webster et al., 2004). Rutherford et al. (2008) found that a 45 s long exposure to a sad face facilitated perceiving a neutral face as happy and vice versa. Leopold et al. (2001) reported that shorter period than 5 s failed to produce after-effects. Interestingly, Fox and Barton (2007) reported identity invariance in facial expression adaptation.

In face studies employing adaptation paradigms, or related paradigms inducing contrast or context effects (Russell & Fehr, 1987; Suzuki & Cavanagh, 1998; Tanaka-Matsumi et al., 1995; Thayer, 1980), the adapting or contextual stimulus typically consists of static face images. However, this is in contrast to natural facial displays, which usually contain dynamic information (cf. Sato, Kochiyama, Yoshikawa, Naito, & Matsumura, 2004), allowing other, additional, processes to play a role in bringing about perceptual biases. One such process is representational momentum (RM), which is generally defined as the phenomenon that an observer's memory for the final position of a moving target is displaced further along the observed trajectory (Freyd & Finke, 1984; Freyd, 1987). The target's motion can either be real, or implied by a sequence of static images. RM results from the observer's inferences regarding the physical dynamics of the (implied) motion (see Hubbard, 2005, for a review). Yoshikawa and Sato (2008) investigated the possible contribution of RM to the perception of dynamic facial expressions by presenting short video-clips depicting a neutral expression morphed into an intense expression. They found that the final expression of the video-clip was overestimated and that the size of this effect was positively correlated with the speed of the intensity change. This bias was explained in terms of Representational Momentum (RM), induced by the preceding sequence.

1.3.3 The influence of explicit and implicit top-down processes in the evaluation of facial expressions

In addition to bottom up visual processes, perceptual evaluations of emotional facial expression can also be influenced by top-down affective or cognitive mechanisms related to the meaning of the facial stimulus. In line with appraisal theories of emotions (Scherer, 1992), it was found that the intensity of angry and fearful faces may depend on eye gaze direction. That is, the intensity of anger increases when the expression is presented with eye gaze directed to the viewer (‘attack detection’), whereas a fearful expression is evaluate as more intense when the eye gaze is averted from the viewer (‘danger detection’: Sander, Grafman, & Zalla, 2003). This finding was interpreted in terms of a cognitive modulation of the perception of facial expressions (Adams & Kleck 2003; Sander, Grandjean, Kaiser, Wehrle, & Scherer, 2007). Similarly, in a recent study it was reported that high-level mental attributions can affect the perception of gaze direction (Teufel et al., 2009). In this study, the agent whose gaze the observers adapted to faced left or right and wore one of two pairs of highly reflective goggles. Gaze adaptation effects increased when participants believed the agent wore the transparent pair compared to when they believed the glasses were opaque. Thus, high level social-cognitive attribution influenced the evaluation of eye gaze direction despite that the bottom-up information was the same in both conditions. This interactive process was referred to by Teufel, Fletcher, & Davis (2010) ‘perceptual mentalising’, which recognizes the crucial influence of mental attributions to low-level visual processes/mechanisms.

However, it is important to note that in these tasks participants’ judgments were guided by explicit attributions. Interestingly, in a recent work it was shown that top-down modulations may be generated automatically and without awareness (Hudson, Liu, & Jellema, 2009; Hudson & Jellema, 2011). In these studies participants’ estimations of how far an agent’s head had rotated were influenced by the agent’s gaze direction. With gaze direction ahead of head rotation the head rotation was overestimated as compared to when the gaze was lagging behind head rotation. This bias seemed to be induced by implicit attributions of the intention to continue/discontinue to move in the direction of the head rotation.

Thus, people’s perceptual evaluation of others’ behaviour (e.g. facial expressions, changes in eye gaze, or head rotations) may not only be affected by visual bottom-up processes, but can also be influenced by top-down processes. These top-down processes can be highly inferential or reflective, but they can also be more reflexive and automatic (Lieberman, 2007; Satpute & Lieberman, 2006). The recognition of the emotion associated to particular facial expression does not involve only a perceptual evaluation but, as described in section 1.2.2.2 above, also motor responses triggered by the observed expression. In the last decade several studies assessed the embodied cognition account on facial expression recognition (Carr et al., 2003; Goldman, 2006; Jabbi, Swart, & Keysers, 2007; Niedenthal, 2007; Pitcher, Garrido, Walsh, & Duchaine, 2008, Wicker et al., 2003).

The discovery of the MNS clarified the role played by the motor and parietal cortices in experiencing the emotional meaning of facial expressions (Rizzolatti & Craighero, 2004; Rizzolatti & Fabbri-Destro, 2008). Therefore, the perception of emotional facial expressions may also be influenced by an automatic and mandatory embodied simulation (Gallese, 2006a, b), which would provide the observer with the meaning of the facial expressions.

1.4 Autism Spectrum Disorders

Autism is currently described as a pervasive developmental disorder, along with Asperger's syndrome, Pervasive Developmental Disorder not otherwise specified, Rett Syndrome and Childhood Disintegrative Disorder (International Classification of Diseases ICD-10, World Health Organization, WHO 1992). Due to the extreme variability in nature and severity of symptoms, autism and Asperger's syndrome are often referred to as Autism Spectrum Disorder (ASD). One idea that has gained recently a lot of support is that each individual in the clinical and typical population occupies a position on this spectrum or continuum, depending on the severity and number of autistic (-like) traits they possess (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). Autism is characterized by impairments in social interaction and communication, and by restricted interests and repetitive behaviour (ICD-10, WHO, 1992; DSM-IV, 2004). Difficulties in social interaction pertain to failures in developing peer relationships, in sharing interests with others and in a lack of emotional reciprocity. Individuals with autism ASD show difficulties in using and comprehending non-verbal behaviour such as eye gaze, facial expressions and body gestures.

Individuals with ASD show deficits in spontaneous language, communication, and non-verbal behavior, which is characterized by rituals and stereotyped or repetitive mechanisms (American Psychiatric Association, APA, 2000).

According to the DSM-IV (2004) a diagnosis of ASD can be made if delays and abnormal functioning are persistent and have had an onset prior to the age of three years and that the disturbances in development cannot be accounted for by a genetic disorder such as Rett's Syndrome (DSM-IV, 2004). Epidemiological data indicates that the incidence of ASD seems to be in a range from 30 to 60 cases per 10000, as compared to the original estimate of 4 per 10000 made 40 years ago (Rutter, 2005). Estimates of the male/female ratio in children with ASD is 4:1 (m:f) (Rutter, 1978; Wing, 1981).

1.4.1 Do individuals with ASD have a Theory of Mind (ToM)?

It has been proposed that individuals with ASD may lack ToM abilities (Baron-Cohen, 1995), as those individuals show difficulties to attribute epistemic mental states (such as beliefs, knowledge, thoughts, desires and intentions) to oneself and others.

Specifically, when trying to interpret others' behaviour, an individual with difficulties in mindreading seems to be thrown back on a temporal regularity account, using routine scripts-based explanations and 'reinforcement schedules'. Baron-Cohen described individuals with ASD as 'mindblind' due to their deficits in the Shared Attention and ToM mechanisms. Several tests have been adopted to examine ToM skills, such as the false belief tasks. It was found that individuals with ASD perform significantly worse than controls on false beliefs tasks (Baron-Cohen et al., 1985; Baron-Cohen, Leslie, & Frith, 1986; Leekam & Perner, 1991; Perner, Frith, Leslie, & Leekam, 1989; Reed & Peterson, 1990; Swettenham, 1996; Swettenham, Baron-Cohen, Gomez, & Walsh, 1996).

However, it has been recognized that in false belief tasks a variety of comprehension problems may interfere with success on these tasks. Interestingly, in a recent study (Senju et al., 2009), children with ASD were tested by adopting the anticipatory looking paradigm that was used by Southgate et al. (2007) on children with typical development. The results revealed that individuals with ASD do not spontaneously attribute mental states to anticipate other people's behavior, as reflected in the absence of anticipatory looking in this task.

Additional investigations focused on the production of spontaneous pretend play in two years-old children with ASD (Baron-Cohen, 1987; Lewis & Boucher, 1988; Ungerer & Sigman, 1981; Wing, Gould, Yeates, & Brierley, 1977). Further, some studies showed that 4-6 years old children with autism have difficulty with mental states as causes of emotion (Baron-Cohen, 1991b; Baron-Cohen, Spitz, & Cross, 1993).

Moreover, it is reported that children with autism show dysfunctional joint attention behaviour, specifically for gaze monitoring (Loveland & Landry, 1986; Mundy, Sigman, Ungerer, & Sherman, 1986; Leekam, Baron-Cohen, Perrett, Milders, & Brown, 1997) and using the pointing gesture (Curcio, 1978; Mundy et al., 1986; Baron-Cohen, 1989).

The ToM hypothesis was criticized because it involves exclusively social deficits while individuals with ASD present impairments also in other domains, such as language, imagination and motor behaviour (Volkmar, Lord, Bailey, Schultz, & Klin, 2004). Moreover, difficulties related to a lack of ToM do not provide a clear explanation for the presence of superior skills in some people with a diagnosis of ASD.

Therefore, others argued that individuals with ASD are motivated to understand others, but due to an inadequate perceptual integration, they fail to apply their ToM skills in complex social situations (Behrmann, Thomas, & Humphreys, 2006; Verbeke et al. 2005).

Other theories emphasize a lack of regulation and control of behavior (Executive Functioning theory; e.g. Russell, 1997), and an aptitude for focusing on parts of objects or small details over global processing (Weak Central Coherence theory; e.g., Happé & Frith, 2006). A related theory, Enhanced perceptual functioning, focuses more on the superiority of locally oriented and perceptual operations in autistic individuals (Mottron, Dawson, Soulières, Hubert, & Burack, 2006).

1.4.2 The 'broken mirror hypothesis' as explanation for the lack of social understanding in autism

Alternatively to the traditional view on the possible causes of social deficits in autism, which ascribes them to a dysfunction in the metalizing network (Frith, 2001; Marsh & Hamilton, 2011), it has been proposed that people with ASD may lack a proper 'mirroring' mechanism. Recent data suggests that a dysfunction of the MNS in humans may lead to core social deficits in autism (Williams, Whiten, Suddendorf, & Perrett, 2001). FMRI studies revealed abnormalities in the inferior frontal gyrus, which is part of the MNS, in ASD (Dapretto et al., 2006).

Additional evidence comes from an EEG study showing that the suppression of the activity of the mu rhythm, which in typically-developed individuals occurs during passive observation of others' actions and is thought to reflect mirroring activity, did not happen in children with autism (Oberman et al., 2005).

These studies are in line with behavioural investigations on individuals with ASD who were found less efficient in processing biological motion (Rutherford, Pennington, & Rogers, 2006). Interestingly, in a recent study it was shown that children with ASD, just like TD children, showed no deficit in intention understanding on the basis of the standard use of an object, while, differently from TD children, they did have difficulties in intention understanding when this relied on motor cues (Boria et al., 2009). Further, children with autism were found to be better at discriminating between pictures of buildings than of faces (Boucher & Lewis, 1992; Boucher, Lewis, & Collins, 1998).

Based on these results, some researchers suggest that core deficits in autism may be linked to impairments of the MNM, causing disabilities in social skills, imitation, empathy and ToM (Gallese & Goldman, 1998; Williams, Moss, Bradshaw, & Mattingley, 2005).

1.4.3 Asperger's syndrome

Within the autism spectrum disorder, some individuals lack verbal impairments but do show difficulties in non-verbal communication, while other deficits appear less extreme as compared to those found in autism. This mild form of autism was first noted by Hans Asperger in 1944, and was later labeled as 'Asperger's syndrome' (Frith, 1991).

Asperger's syndrome (AS) is a type of pervasive developmental disorder, included in the ICD-10 (World Health Organisation, WHO, 1992) and the DSM-IV-TR (APA, 2000). Individuals with AS have an intelligence that is within the normal IQ range or above, and do not manifest a significant delay in language development.

However, they do show deficits in reciprocal social interaction, unusual verbal and non-verbal communication and the presence of idiosyncratic isolated interests.

Recently a review of the diagnosis of Asperger's syndrome has been proposed with a possible new nomenclature in the next DSM-V (2013). Autism and Asperger's syndrome may be replaced by a more concise diagnostic criteria of 'Autism Spectrum Disorder' in which several 'tiers' of severity will be adopted (DSM-V). Some clinicians suggested the use of 'high functioning autism' instead of Asperger's syndrome, or in addition to Asperger's syndrome (Venter, Lord, & Schopler, 1992, Rumsey & Hamburger, 1988). However it is not clear whether high-functioning refers to verbal, or non-verbal, intelligence, or perhaps to relatively high social adaptation regardless of intelligence (Frith, 1996).

High functioning autism (HFA) presents symptoms that overlap with both autism and Asperger's syndrome (Macintosh & Dissanayke, 2004). Individuals with HFA are by definition more cognitively able than those with autism, but they still show motor and language impairments, which are absent in individuals with Asperger's syndrome. Therefore, HFA and AS individuals differ not only in terms of cognitive functioning, but also in symptomology and early developmental history (Miller & Ozonoff, 2000). Specifically, individuals with Asperger's syndrome manifest less severe early symptoms, a milder developmental course and a better prognosis, than individuals with HFA. Overall, Asperger's syndrome and HFA present a different degree or severity of the same symptoms.

1.4.3.1 The processing of social cues and action anticipation in individuals with Asperger's syndrome (AS)

It has been reported that individuals with AS find it difficult to process nonverbal social cues, such as eye contact, facial expression, and body gesture. Such difficulties may be related to a failure to attribute epistemic mental states to others (Frith, Morton, & Leslie, 1991; Baron-Cohen, 1995; Baron-Cohen, Campbell, Karmiloff-Smith, Grant, & Walker, 1995) and to spontaneously interpret others' social cues (Senju, Southgate, White, & Frith, 2009; Jellema et al., 2009). Individuals with AS may find difficult to empathize with others, showing a lack of social or emotional reciprocity (McPartland & Klin, 2006). However, despite individuals with AS having difficulty in understanding others' mental states spontaneously, they may rely on statistical regularities or input-output relationships which typically govern the motions of non-agentive objects to explain human actions, to compensate for their failure in social understanding (Kuchinke, Scheider, Kotz, & Jacobs, 2011; Kuzmanovic, Schilbach, Lehnhardt, Bente, & Vogeley, 2011; Senju et al., 2009).

One of the social cues which could be problematic to decipher by AS individuals is the expressivity of faces. A recent review on this topic in autism revealed mixed results (Harms, Martin, & Wallace, 2010). This is mainly due to methodological issues, such as the use of a variety of paradigms and stimuli (usually static pictures which are less ecologically valid), different age groups and different levels of severity of ASD.

Therefore, whereas some studies showed impairments in emotional facial processing (e.g. Celani, Battacchi, & Arcidiacono, 1999; Hobson, 1986a, 1986b; Law Smith, Montagne, Perrett, Gill, & Gallagher, 2010; Lindner & Rosen, 2006; Tantam, Monaghan, Nicholson, & Stirling, 1989; Teunisse & de Gelder, 2001), other studies reported no impairments (e.g. Adolphs, Sears, & Piven, 2001; Baron-Cohen et al., 1997; Castelli, 2005; Gepner, Deruelle, & Grynfeltt, 2001; Loveland et al., 1997; Ogai et al., 2003; Ozonoff, Pennington, & Rogers, 1990; Prior, Dahlstrom, & Squires, 1990; Wicker et al., 2008; Wong, Fung, Chua, & McAlonan, 2008).

However, studies which were specifically focused on individuals with AS or HFA mostly revealed that while the processing of basic emotions is relatively preserved, subtle or ambiguous facial expressions such as arrogance or flirtatiousness do pose problems (e.g. Baron-Cohen et al., 1997; Kleinman, Marciano, & Ault, 2001). Another study with eye tracking revealed that facial expression recognition depend on whether participants fixate the eye region, which is extremely informative when the face signals fear or anger (Corden, Chilvers, & Skuse, 2008). It should be noted that almost all of these studies were conducted with static facial displays. Interestingly, when dynamic stimuli were employed facial affect recognition was relatively improved (Gepner et al., 2001; Loveland et al., 1997).

With regard to ToM skills, individuals with AS tend to pass false belief tasks (Abell, Happe, & Frith, 2000; Baron-Cohen et al., 1997). However, they still report difficulties in attributing mental states to geometric shapes on the basis of their movement pattern (Abell et al., 2000), or in attributing mental states to other people from looking at photographs of their eyes (Baron-Cohen et al., 1997; Baron-Cohen et al., 2001).

Advanced tests for impairments in the ToM domain focused on the ability of individuals with AS to *anticipate* others' actions on the basis of social cues (Zalla, Labruyere, Clement, & Georgieff, 2010; Hudson, Burnett, & Jellema, 2011). Zalla et al. (2010) showed videotaped movies of an actor performing familiar and non-familiar actions. The authors found that children and adolescent with AS were less correct in predicting the outcome of the goal-directed action even when it was a familiar action, as compared to the control groups. However, recently it has been suggested that these deficits may be compensated by alternative strategies, such as relying on physical dynamic characteristics of the agent (Hudson et al., 2011). In this study by Hudson et al. (2011) TD and AS adults needed to evaluate how far an agent's head had rotated in conditions where the eye gaze was leading the direction of the rotation or was lagging behind head rotation. Participants observed faces but also an inanimate cylindrical object containing the equivalent of eyes but not resembling a face at all. It was found that both TD and AS individuals believed that the agent's head had rotated further when the agent's gaze was ahead of rotation than when it was lagging behind rotation. However, while in TD individuals such anticipatory behaviour was only found for faces and not for the cylindrical objects, in AS individuals the anticipatory behaviour was found in both the animate and inanimate conditions. This suggested that while the evaluations of TD individuals were affected by the involuntary attribution of intentions to the agent, in the AS group the action anticipation was caused by the low-level visual parameters (present in both animate and inanimate conditions) rather than by the inferred intentions.

To summarize, social cue processing is considered to play a crucial role in the involuntary understanding of other people's intentions underlying their actions, and may also lead to an anticipation of other's actions in the immediate future. Emotional facial expressions are thought to be immediate signifiers of the affective dispositions of others. Theory-Theory claims that people understand someone else's emotional state by positioning it inside a theoretical framework. Alternatively, Simulation Theory suggests people grasp others' emotions by simulating 'pretend states' in themselves.

According to the 'embodied simulation' hypothesis (Gallese, 2006a, b), these simulations can be understood as involuntary activations of one's own motor system, employed as a template to model the experiences of others. The generated experience reveals to the observer the meaning of the observed expressions. When looking at someone else's emotional facial expression, what the observer sees are not only literal facial characteristics (features and configurations) and a variety of different motions, but also the agent's emotional state of mind. It is as if we directly perceive the emotional/mental state *in* the facial expression, a point emphasised by the phenomenological tradition in philosophy of mind (Merleau-Ponty, 1964).

Previous research showed that the evaluation of facial expressions can be influenced by both bottom-up low-level visual processes and by *explicit* inferences made about others' mental states. However, in a recent study it was found that *implicit* attributions made by the observer about the agent's intentions can also influence quite early stages of the visual processing of social stimuli (Hudson et al., 2009). When facial expressions are embedded in a dynamic action sequence the immediately preceding perceptual history may generate expectations in the observer. This may lead one to anticipate how the emotional state of the agent would develop in the immediate future. Therefore, implicit understanding or anticipation of others' affective states in turn can bias (top-down) the evaluation of the displayed facial expression.

Mind-reading skills (Baron-Cohen, 1995) are dysfunctional in individuals with autistic spectrum disorders (ASD) Baron-Cohen et al., 1991b; Frith, 2003; Frith et al., 1991). Individuals with ASD may analyze facial expressions on the basis of isolated elements, or low-level features (e.g. the U shaped mouth in a smile), missing the associated emotional meanings. Individuals with Asperger's syndrome (AS), a mild form of ASD, may be able to metalize but fail to do so spontaneously (Jellema et al., 2009; Kuzmanovic et al., 2011; Senju et al., 2009), which determines socially inadequate behaviour. The lack of the ability in individuals with AS to implicitly attribute emotional states to others may be reflected in the evaluation of dynamic facial expressions.

1.5 Aims of the thesis

The current thesis present a series of studies aimed at examining the relative contributions of low level visual mechanisms and of implicit attributions about an agent's emotional states on perceptual judgments of the agent's dynamic facial expressions, in adults with TD and AS. Further investigations tried to manipulate the intention behind the displayed facial expressions to test whether this could modulate the evaluations of the expressions in question.

Research questions:

(1) When TD individuals are presented with dynamic sequences of emotional facial expressions, is their evaluation of the end-point of those sequences influenced by low-level visual mechanisms? Specifically, what is the contribution of (1) contrast or context effects, (2) adaptation and (3) extrapolation or Representational momentum (RM)? In principle, each of these mechanisms could produce a bottom-up distortion of the perception of the emotional facial expressions.

(2) Alternatively, the ecologically-valid dynamic action sequence ('perceptual history') in which the emotional facial expressions are embedded may give rise to an anticipation of how the emotional state of mind will develop in the immediate future. This 'emotional anticipation' mechanism (emotion reading; Goldman, 2006) may bias (top-down) the visual perception of the current facial expression in the direction of the anticipated state of mind. In this case, the observers' evaluations would be influenced by an implicit attribution made about the affective state of the agent. To answer these questions, in Chapter 2 a series of six experiments were conducted to

disentangle the contributions of those mechanisms to the perception of dynamic emotional facial expressions in TD individuals.

(3) How would individuals with Asperger's syndrome (AS) evaluate the dynamic emotional facial expressions? What would be the relative contributions of the low-level visual mechanisms and the emotional anticipation in this group? Which kind of mechanisms may account for possible perceptual bias when judging the end-state of emotional facial expressions? Due to the difficulty in emotion reading eventual perceptual distortions may depend on a mechanistic analysis based on the physical proprieties of the stimulus rather than on the emotional information it conveys (Study 2, Chapter 3). To answer these questions, in Chapter 3 a series of four experiments were conducted to examine whether the same mechanisms would be active in individuals with AS. If the results in individuals with AS on the tasks would be different from TD individuals, then it will be determined which mechanisms are used (and which not).

(4) Which kind of mechanism would be triggered if the dynamic facial expressions are not linked to a genuine emotional state of mind? Does a dynamic sequence of 'fake' facial expression generate an automatic emotional anticipation? How does explicit information about the meaning of the facial expressions guide participants' evaluations? To answer these questions, in Chapter 4 the emotional anticipation hypothesis was tested in TD individuals using deceptive, or 'fake', facial expressions displayed by professional actors.

(5) What are the neural mechanisms responsible for the overshoot response bias? Is emotional anticipation underpinned by embodied simulation of the observed

facial expressions? To make a start with answering these questions, an electromyographical study was performed (Appendix), in which activity in the Zygomaticus Major and Currugator Supercilii was recorded from participants while they observed the dynamic facial expressions.

CHAPTER 2

Emotional anticipation in adults with typical development

2.1 Introduction

It is virtually impossible to look at an expression displayed on a human face and not automatically get a notion about the individual's emotional state of mind. It is as if we directly perceive the emotional/mental state *in* the facial expression, a point emphasised by the phenomenological tradition in philosophy of mind (e.g. Merleau-Ponty, 1964). Wittgenstein, although not a phenomenologist, remarked: "We *see* emotion". - As opposed to what? - We do not see facial contortions and *make the inference* that he is feeling joy, grief, boredom. We describe a face immediately as sad, radiant, bored, even when we are unable to give any other description of the features. - Grief, one would like to say, is personified in the face." (Wittgenstein, 1980, Vol. 2, Section 570). These ideas have been reinvigorated by contemporary philosophers (Gallagher & Zahavi, 2008), and were echoed by a renewed interest into the interplay between bottom-up visual processes and top-down attributions of mental/emotional states or Theory of Mind (cf. Teufel et al., 2010).

New evidence emphasises that quite early stages of the visual processing of social stimuli, such as the processing of other's physical postures, which were previously thought to be fairly autonomous and immune to top-down influences, can be influenced by (implicit) attributions made by the observer about the agent's mental/emotional state of mind. For example, Hudson et al. (2009) and Hudson & Jellema (2011) found that estimations of how far an agent's head had rotated were influenced by the agent's gaze direction. With gaze directed ahead of head rotation, the angle of head rotation was overestimated as compared to when the gaze was lagging behind head rotation. This bias seems to be induced by implicit attributions of the intention to continue/discontinue to move in the direction of the head rotation. Further, it was demonstrated using a gaze-adaptation paradigm that high-level mental attributions of ‘seeing’ versus ‘not seeing’ affect the basic mechanism of gaze perception differently (Teufel et al. (2009). In this study, the agent whose gaze the observers adapted to faced left or right and wore one of two pairs of highly reflective goggles. When observers believed the agent wore the transparent pair, the gaze adaptation effect increased compared to when the observer believed the glasses were opaque. Thus, attribution of a ‘seeing’ mental state increased gaze processing relative to attribution of a ‘non-seeing’ mental state, despite identical perceptual bottom-up information in both conditions.

It should be noted, though, that while in the Teufel et al. (2009) study the mental attributions reflected explicit knowledge, in the Hudson et al. (2009) study the attribution of the mental states occurred implicitly, as participants did not recall having seen the gaze manipulation. The studies presented in Chapter 2 investigated on implicit attribution/anticipation made by the observer about the agent’s emotional state of mind and how these may influence the perception of dynamic facial expressions.

Thus, whereas previously it had been argued that the bottom-up processing stages involved in the perception of an agent give rise to inferences about the mental/emotional state of mind of the agent, which are then subsequently attributed to the agent (e.g. Blakemore & Decety, 2001), it is now becoming clear that these emotional/mental attributions already affect the very basic processing stages in social perception. This highlights the bi-directional interaction between bottom-up and top-down streams, which has been captured under the term ‘perceptual mentalizing’ (Teufel et al., 2010).

2.1.1 Influence of mental attribution on the perception of facial expressions

The study presented in Chapter 2 aimed at examining to what extent the perception of low-level facial characteristics, such as the ‘neutrality’ of facial expressions, is affected in an automatic manner by mental/emotional attributions to the agent on the basis of the immediate perceptual history. It utilized the idea that experimentally-induced perceptual distortions can offer a handle on the investigation of the relative contributions of bottom-up and top-down processes.

A new perceptual distortion of neutral facial expressions induced by the immediately preceding perceptual history of the face was previously described (Jellema et al., 2011). This study showed that neutral facial expressions were judged as slightly angry when they were immediately preceded by video-sequences that depicted a facial expression of joy that gradually morphed into a neutral expression, while the identical neutral expressions were judged as slightly happy when the videos started off with an angry expression that morphed into a neutral one. Thus, the participants showed a perceptual ‘overshoot’ response bias.

Participants were instructed to evaluate the last expression of the Joy-to-neutral and Anger-to-neutral video-clips on a 5-point scale ranging from slightly angry via neutral to slightly happy. Thus, even though the last frames depicted identical neutral expressions, they were perceived differently depending on the specific perceptual history. The video-sequences were ecologically valid, meaning that the morphed Joy-to-neutral and Anger-to-neutral sequences resembled movements occurring during naturalistic dynamic facial expressions (Sato & Yoshikawa, 2004). Jellema et al. (2011) employed these video-sequences in an affective priming paradigm as task-irrelevant distractors, while the targets were formed by positive or negative words superimposed on the last frame of the sequences. Participants had to make speeded evaluations of the word valence. The study found that positive words were faster detected when superimposed on the Anger-to-neutral videos (as compared to Joy-to-neutral videos), and negative words were faster detected when superimposed on the Joy-to-neutral videos (as compared to Anger-to-neutral videos). The important conclusion from this experiment was that the Joy-to-neutral sequence induced a negative affect, and the Anger-to-neutral sequence a positive affect, in the observer. It further suggested that the perceptual 'overshoot' biases were not due to a cognitive response strategy, such as selecting the opposite of the start-emotion, but reflected a genuine change in the observer's judgement. A speculative interpretation of the results of the Jellema et al. (2011) study was that the perceptual distortions were due to the observer involuntarily keeping track of the emotional state of mind of the actor, leading to an involuntarily anticipation of the most likely future emotional state. The anticipated emotional state would then in turn affect the perception. Thus, here the context affecting perception was provided by dynamic information contained in the social stimulus itself, rather than by external information provided by the experimenter, as in Teufel et al. (2009). The underlying process was called 'emotional anticipation', referring to the ability to anticipate how the agent's emotional state of mind will develop in the immediate future, based on the immediate perceptual history. This could be seen as a kind of low-level mindreading (cf. Goldman, 2006). That is, in the Joy-to-neutral condition the observer expects the agent's emotional state to continue to move into a 'negative' state, as a consequence of a decrease of positive affect and in the Anger-to-neutral condition to continue to move into a 'positive' state as a consequence of a decrease of negative affect. The emotional anticipation hypothesis is also related to the contention that humans with typical development have a Theory of Mind (ToM), meaning they are equipped with the ability to ‘read’ others’ mental states, such as beliefs, desires, intentions and emotions (Baron-Cohen, 1995).

Why would observers attribute emotional/mental states to the agent in the Joy-to-neutral and Anger-to-neutral conditions?

The ability to predict how the emotional state of mind of individuals may change across time (whether in a positive or negative sense) is crucial for successful social interactions, and also opens up possibilities to persuade others (Weiner & Handel, 1985; Petty & Cacioppo, 1986). With respect to the specific emotional sequence presented in the Jellema et al. (2011) study, it seems fair to assume that a naturally occurring change of relatively short duration (a few hundred ms, e.g. Sato, Kochiyama, Yoshikawa, Naito, & Matsumura, 2004; Yoshikawa & Sato, 2008; Sato & Yoshikawa, 2010) from a maximally happy expression into a neutral expression constitutes a negative signal, and that the most likely anticipation of the agent’s future emotional state of mind would be a negative emotional state. Thus the trend in the emotional change is anticipated to continue, which inevitably leads to a negative emotional state of mind. Similarly, a sudden change from a maximally angry expression into a neutral expression constitutes a positive signal, and the agent’s emotional state of mind would be anticipated to continue to move into a positive emotional state of mind.

However, the emotional anticipation hypothesis, which is still highly speculative at this stage, is not the only plausible candidate to explain the overshoot response bias found in Jellema et al. (2011). Basic low-level visual mechanisms might just as well have contributed to the overshoot response bias.

2.1.2 The influence of basic perceptual processes

A number of bottom-up visual processes could have caused, or contributed to, the distorted perception of the neutral expression in the above paradigm. These include (i) (sequential) contrast/context effects, (ii) extrapolation or representational momentum (RM), and (iii) adaptation.

(i) Contrast/context effects.

In contrast/context effects, the target stimulus is influenced by the presentation of an inducing stimulus or anchor (e.g. Russell & Fehr, 1987; Suzuki & Cavanagh, 1998; Tanaka-Matsumi et al., 1995; Thayer, 1980). Thayer (1980) showed that intensity ratings of both sad and happy facial expressions were enhanced when they were preceded by contrasting, as opposed to similar, facial expressions. Russell & Fehr (1987) presented two facial expressions of different identities side by side and found that the judgment of the first face (anchor) affected the judgment of the second face (target) in a repulsive manner. The results of this study were, however, criticized by Ekman & O'Sullivan (1988), who maintained that the perception of emotion in facial expressions should be accurate and absolute, though they agreed that neutral or ambiguous expressions may be subject to contrast/context effects. Tanaka-Matsumi et al. (1995) found that a neutral face (the target) was rated as less pleasant and less arousing when preceded by a happy face (the anchor) as compared to when preceded by a sad anchor face. A common finding in these contrast/context effect studies was that it did not matter whether the identities of the actors depicting the emotion in the anchor and target stimuli were the same or different, the effect would always be there (Thayer, 1980). What the studies methodologically had in common was that the anchor and target stimuli formed a less ecologically valid sequence, which consisted of discrete stimuli conveying an emotion, as compared to a gradual transition from one expression into the other for one and the same identity. Nevertheless, in the Jellema et al. (2011) study the contrast between the initial maximally happy or angry expression and the final neutral expression may have made the last neutral expression look slightly angry or happy, respectively.

(ii) Representational momentum (RM).

RM refers to the finding that the observer's memory for the final position of a moving target is displaced further along the observed trajectory. It results from the observer's inferences regarding the physical dynamics of the movement, and is hypothesized to reflect an anticipatory function (Finke & Freyd, 1985; Freyd & Finke, 1984). The extent of the memory displacement depends on the physical causes (e.g. gravity) and constraints (e.g. friction) that are inferred to act upon the object's motion (Hubbard, 1995). The displacement even occurs for static images of implied motion (Freyd, 1983) and can be influenced by conceptual knowledge the observer has about the nature of the object (Vinson & Reed, 2002). In the study by Jellema et al. (2011) RM might also act on the gradual dynamic changes of the facial features, such as the U-shaped mouth in the happy expression, which morphs into a flat shape in the neutral expression and might then be extrapolated into an inverted U shape, giving the impression of a slightly angry face. However, it could, in principle, also act on an underlying positive-to-negative valence dimension. Recently, Yoshikawa and Sato (2008) presented short video-clips depicting neutral expressions that morphed into intense emotional expressions, while observers had to judge the intensity of the expression depicted on the last frame. They found that the intensity of this final expression was overestimated and that the size of this effect was positively correlated with the speed of the intensity change. They explained the bias in terms of an RM effect, induced by the preceding sequence. They did, however, not specify which dimension the RM effect was supposed to operate on.

(iii) Adaptation.

Adaptation, which is ubiquitous in visual processing, is another candidate for explaining the overshoot response bias. A prolonged observation of a distorted face causes a normal (test) face to look somewhat distorted in the opposite direction. Face after-effects induced by adaptation were found for emotional expressions (Fox & Barton, 2007; Hsu & Young, 2004; Rutherford et al., 2008; Webster et al., 2004) and for a variety of other facial traits.

2.1.3. Aim of the current study

The studies presented in Chapter 2 were conducted to investigate the causes of the perceptual distortion in the above described paradigm (Jellema et al., 2011), by trying to experimentally disentangle the relative contributions of basic bottom-up perceptual processes from possible top-down attributional processes. The Jellema et al. (2011) study merely speculated about possible causes for the overshoot bias, no attempt were made to investigate the underlying mechanism. The dynamic perceptual histories used in the current study, depicting intense emotional expressions morphing into neutral expressions, were thus effectively the reverse of those presented by Sato & Yoshikawa (2008). They were ecologically more valid (Sato & Yoshikawa, 2004), and may therefore produce different effects, than perceptual histories consisting of static facial expressions as typically used in adaptation and contrast-effect studies.

Six experiments were conducted to test whether the overshoot bias may be due to basic visual contrast effect, adaptation, extrapolation or RM, or whether it would be better explained by top-down emotional anticipation.

In Experiment 1 the endpoint of the clips was varied to test whether a crossing of the category-boundary would occur even when the endpoint belonged to the same category as the starting point. In Experiment 2 the identity of the agent was changed at the end of the clips, which should not affect contrast/context effects, RM on the underlying positive-negative valence dimension, nor adaptation, but would affect emotional anticipation. In Experiment 3 the contribution of contrast/context effects was further examined by employing sequences which started from a neutral expression and morphed via happy (or angry) back to the same neutral expression. In Experiment 4 the speed of the video-clips was manipulated to further examine the role played by RM. In Experiment 5 a blank was inserted just before the final neutral expression to investigate whether disruption of the flow of motion present in the video-sequence would affect the evaluation due to a reset of low-level visual mechanisms (cf. Enns, Lleras, & Moore, 2010). Finally, in Experiment 6 the duration of the first frame of the video-sequence was doubled to further test the possible contribution of visual adaptation.

2.2 Experiment 1

Experiment 1 aimed at assessing whether the crossing of the category boundary in the perceptual evaluation of the neutral expression depicted in the last frame of the videos, (Jellema et al., 2011), would even occur when the video's endpoint depicted an emotion of the same category as the emotion at the start of the video. Short video-clips were presented of dynamic facial expressions that morphed from an expression of intense joy or anger to a neutral expression or to a 10% joy or a 10% anger expression. If the overshoot would occur also when the end-point is not neutral but belongs to the same emotion category as the starting point, then this would reveal the strength of such effect.

Participants were required to evaluate the last expression of the sequence.

2.2.1 Method

2.2.1.1 Participants

Thirty-nine undergraduate Psychology students at Hull University (UK) took part in the experiment. All participants had normal or corrected-to-normal vision, gave informed consent, and received course credit for taking part. After applying exclusion criteria (see data reduction below), the data of thirty-three participants were included in the analysis (age, M = 20.1 years, SD = 3.2 years; 28 females, 5 males).

2.2.1.2 Stimuli

Pictures of eight actors displaying facial expressions of joy and anger were selected from the Pictures of Facial Affect (four males: EM, JJ, PE, WF, and four females: C, MO, PF, SW; Ekman & Friesen, 1976).

All faces were shown from frontal view with their eye gaze directed straight ahead. The photographs were in greyscale. The hair had been blackened so as to merge with the black background. The pictures were digitally adjusted to match in contrast and brightness. The eyes of all actors were positioned on approximately the same location on the screen. Faces measured about 13 x 20 cm when displayed on the screen, subtending roughly 8° vertically.

Nine interpolated images, in between the full-blown expression of joy or anger (which is called 100%) and the neutral expression (0%) were created at equal steps of 10% intensity change, using computer morphing procedures (Perrett, May, & Yoshikawa, 1994). Rapid successive presentation of these interpolated frames constituted the videos. The first frame of each video sequence showed the emotional expression at 100% intensity and was presented for 300 ms to ensure the type of emotion was properly recognised (Eimer & Holmes, 2007; Palermo & Rhodes, 2007). The subsequent interpolated frames were shown for 30 ms each, as reported in previous studies which employed similar video-clips (Sato & Yoshikawa, 2010). The clips ended either in a neutral expression, in a 10% joy expression or in a 10% anger expression. The last frame remained on the screen for 300 ms as the starting frame, to control temporal effects. The total duration of the morph sequence was 270 ms (9 x 30 ms), and was 30 ms longer or shorter for clips with 10% intensity endpoints (see Figure 2.1). A total of 48 videos were made (8 actors x 2 perceptual histories x 3 endpoints).

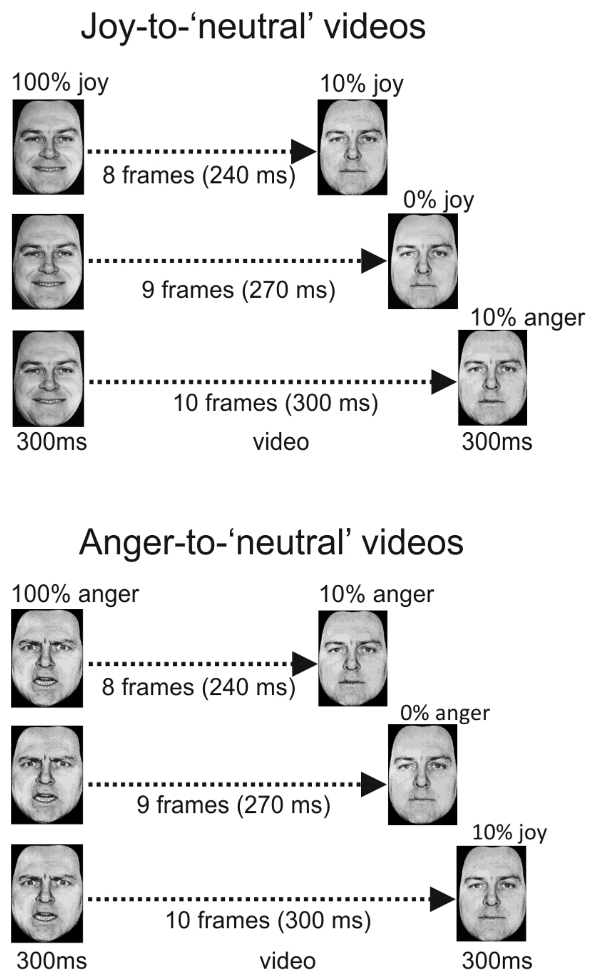

Figure 2.1. Schematic illustration of the stimulus presentations in Experiment 1. Joy-to-‘neutral’ videos started with a facial expression of intense joy (100%), which gradually morphed into a 10% joy, a neutral or a 10% anger expression (top panel). Anger-to-‘neutral’ videos started with a facial expression of intense anger (100%), which gradually morphed into a 10% anger, a neutral or a 10% joy expression (bottom panel).

2.2.1.3 Measurement of the perceptual evaluations: The response scale

As in Jellema et al. (2011), the participants’ perceptual judgments of the facial expressions were measured on a scale where *verbal* labels indicated the different facial expressions and their intensities (from slightly angry to neutral to slightly happy), rather than a scale formed by the facial expressions themselves. Providing the participant with different pictures of facial expressions and requiring them to select the expression that most resembled the expression at the end of the clip (cf. Yoshikawa & Sato, 2008) would be problematic for the current experimental purposes. The process of going through the range of pictures of emotional expressions (e.g. by using an interactive slider) would in itself, unwillingly, create a new perceptual history. This new perceptual history would than interfere with the crucial perceptual history, which immediately precedes the to-be-judged final neutral expression. Also a procedure by which the participant selects, out of two different expressions, the expression that best resembled the clip's end-expression would not be appropriate, as it could induce contrast effects with respect to the end-expression.

Why 'joy' and 'anger', and not 'joy' and 'sadness'?

Expressions of joy and anger were selected as the two end-points on the 5-point emotion response scale, rather than joy and sadness, on the basis of the following considerations. The end-points on the scale should ideally be equal in terms of approach/avoidance, arousal, and the extent to which their distinctive physical features differ geometrically from those in the neutral expression. This is important as it allows to exclude that a possible response bias on the 5-point scale is due to any of these factors. Joy and anger are both approach-oriented emotions, whereas sadness is an avoidance-related emotion (cf. Davidson & Hughdahl, 1995). Therefore, differences in responses could not be related to different positions on the underlying approach/avoidance dimension. According to the Circumplex model of emotion (Russell, 1980; Russell & Bullock, 1985), in which facial expressions are represented by values on the continuous dimensions of arousal and pleasantness, joy and anger are similar in terms of arousal (whereas joy and sadness are not). Ideally, the extent of geometrical change in the face when morphing from 100% intensity to neutral should be identical for the different emotional expressions used, to exclude the possibility that response biases are due to differences in the amount of change in physical features, which could cause e.g. differences in the extent of extrapolation of these curves. Joy and anger are in this respect fairly similar, whereas happy and sad are not (as the sad expression is less expressive than the happy and angry ones (cf. Calvo & Marrero, 2009). This is also reflected by the finding that the expressions of joy and anger are both detected faster and more accurately in visual search tasks than the expressions of sadness and fear (Williams et al., 2005). The above findings render the sad expression unsuitable for use on the 5-point scale opposite to joy. However, according to the Circumplex model, joy and anger do differ maximally in terms of valence, as reflected by their positions on the pleasantness dimension, which is an essential requirement for the current experiment.

Other facial expressions were considered inappropriate for the ‘overshoot’ paradigm for various reasons. The surprise expression was not appropriate because it has a twofold emotional valence (Neta & Whalen, 2010), while disgust was not used because it less recognizable, as it may present a mixture of other basic emotions (Kohler et al., 2003). Allowing participants to choose from all six basic emotions (joy, surprise, anger, sadness, disgust and fear; Ekman & Friesen, 1997) might induce a bias towards negative responses as the majority of basic emotions is negative.

It should be noted that this scale was not a forced two-choice scale (angry or happy) as participants could select the middle point of the scale, corresponding to a neutral expression. Nevertheless, it might that the particular labels 'anger' and 'joy' were somehow instrumental in producing the 'overshoot' response bias. To examine this, a pilot study was conducted on 8 participants (age, M = 20.4 years, SD = 2.5 years; 7 females, 1 male), in which the labels 'anger' and 'joy' were replaced by the labels 'negative' and 'positive'. Thus the scale in this pilot ranged from slightly negative (1) via neutral (3) to slightly positive (5). The results showed a robust overshoot effect, similar to that obtained using the 'anger' and 'joy' labels: the expression in the last frame was perceived as more negative in the Joy-to-neutral videos (M = 2.65, SD = 0.33) as compared to the Anger-to-neutral videos (M = 3.46, SD = 0.54; $t(7) = -4.86$, $p = .002$).

2.2.1.4 Experimental procedure

Participants were seated at a viewing distance of 80 cm from a PC screen (17-inch monitor, 1024 x 768 pixels, 100 Hz). The stimuli were presented using E-Prime (v. 1.2; Psychology Software Tools, Inc.).

First a calibration condition was presented in which the eight actors were shown with static neutral expressions (i.e. neutral expressions according to the ratings from Ekman & Friesen, 1976). Each calibration trial started with a fixation cross displayed in the centre of the screen for 500 ms, followed by a single frame depicting the neutral face, presented for 600 ms. Sixteen calibration trials were presented (8 actors, 2 repetitions each) in random order. Participants were prompted to rate these 'neutral' expressions using the 5-point scale, ranging from slightly angry (1) via neutral (3) to slightly happy (5), by pressing one of the 5 labeled keys on a button box (SR-Box, Psychology Software Tools, Inc., USA).

After the calibration trials, 8 practice trials were completed, followed by 64 randomised experimental trials, half of which started with a happy expression (called joy-to-'neutral' videos) and half with an angry expression (called anger-to-'neutral' videos). Of both types of videos, 50% of trials ended with the neutral expression, 25% of trials ended with the 10% joy and 25% of trials with the 10% anger expression. Each trial started with a fixation cross (500 ms), followed by the video. As soon as the video ended a blank screen appeared with a prompt to provide a response. The participant's task was to evaluate the last expression using the same 5-point scale as for the calibration trials. Participants were instructed that their RTs were irrelevant, but that they should give their response within 3 seconds. The duration of the entire experiment was 20 minutes.

2.2.1.5 Data reduction and analysis

Trials in which the RTs fell below 250 ms or above 3000 ms were considered outliers and were removed (4.8%). Participants were excluded if more than 25% of the RT values were outside the 250 – 3000 ms range (n = 4). Further a $\pm$2.5 SD rule was applied to the mean difference ratings (mean rating in Anger-to-neutral condition minus mean rating in Joy-to-neutral condition), which excluded 2 more participants.

The calibration scores were used to adjust the scores in the experimental trials; a calibration factor of [3.00 minus the calibration score] was added to the raw scores. For illustrative purposes the calibration scores for the neutral expression of each actor as obtained in Experiment 1 (n = 33) are shown in Figure 2.2. A very similar pattern of calibration scores was obtained across experiments, with the neutral expression of actors C and WF systematically rated as slightly angry. The calibration procedure allowed to perform one-sample t tests with test value 3.00 for each perceptual histories. All statistical analyses were performed on the calibrated scores.

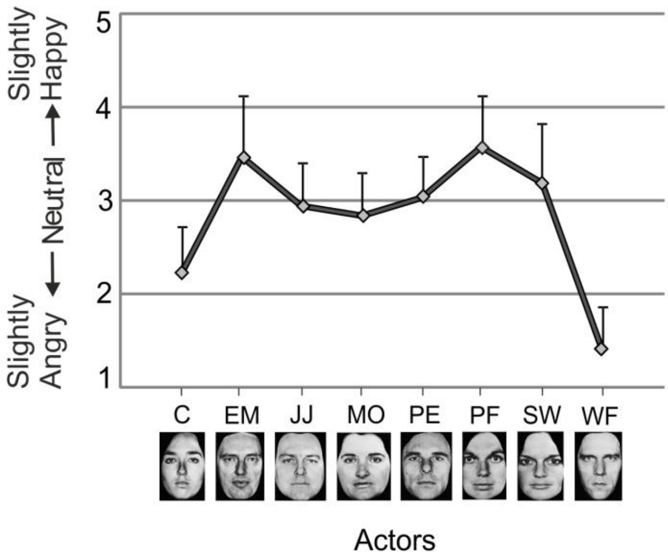

Figure 2.2. Calibration scores in Experiment 1. Ratings on the 5-point scale (vertical axis) for the neutral expressions of the 8 actors presented in the calibration phase. Error bars indicate +1SD.

2.2.2 Results

The results are shown in Figure 2.3. A 2x3 repeated measures ANOVA with Perceptual history (joy-to-‘neutral’ videos vs. anger-to-‘neutral’ videos) and Endpoint (10% joy vs. neutral vs. 10% anger) as within-subject factors, showed a robust significant main effect for Perceptual history ($F(1, 32) = 67.4$, $p < .0001$, $\eta_p^2 = .68$). The expression in the last frame was perceived as more angry in the joy videos (M = 2.81, SD = 0.34) as compared to the anger videos (M = 3.43, SD = 0.35), which reflects the ‘overshoot’ response bias. There was also a significant main effect for the factor Endpoint ($F(2, 64) = 58.9$, $p < .0001$, $\eta_p^2 = .65$), with expressions in the 10% anger endpoint condition evaluated as more negative than those in the neutral endpoint condition ($t(32)= -3.32$, $p = .002$), and expressions in the neutral endpoint condition as more negative than those in the 10% joy endpoint condition ($t(32) = -8.05$, $p < .001$). The Endpoint by Perceptual history interaction was significant ($F(2, 64) = 3.77$, $p = .028$, $\eta_p^2 = .11$). In the neutral endpoint condition, the mean ratings in the joy-videos (M = 2.76, SD = .30) and anger-videos (M = 3.34, SD = .33) differed significantly from each other ($t(32) = -6.45$, $p < .0001$), and each of them differed significantly from 3.00 (joy-video: $t(32) = -4.67$, $p < .0001$; anger-video: $t(32) = 5.94$, $p < .0001$; two-tailed).

In the 10% joy endpoint condition the ratings in the two perceptual histories were M = 3.00, SD = .35 for the happy-videos and M = 3.76, SD = .38 for the anger-videos, while in the 10% anger endpoint condition they were M = 2.66, SD = .37 for the happy-videos and M = 3.20, SD = .34 for the anger-videos.

Remarkably, the 10% joy endpoint in the joy-videos was rated as significantly more angry than the 10% anger endpoint in the anger-videos ($t(32) = 2.34$, $p = .025$), despite the former expression being 20% more ‘positive’ than the latter.

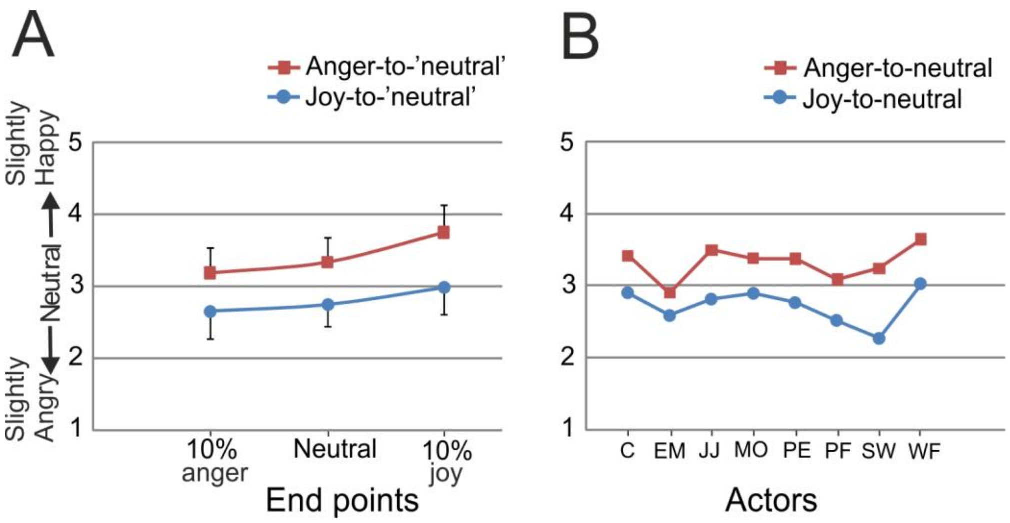

Figure 2.3. Results of Experiment 1. (A) Ratings on the 5-point scale (vertical axis) for the expressions depicted on the last frame of the joy-to-‘neutral’ and anger-to-‘neutral’ videos. The sequences ended at 10% anger, neutral or 10% joy (horizontal axis). Error bars indicate 1SD. (B) Ratings for exclusively the neutral expressions at the end of the happy- and angry-videos. Scores are shown for each of the 8 actors (horizontal axis) to illustrate the consistency of the bias across actors.

2.2.3 Discussion

In Experiment 1 videos starting with joy or anger ended either with a neutral, a 10% joy or a 10% anger expression. This manipulation allowed to test whether the overshoot bias would occur even when the end-point of the videos depicted an expression belonging to the same emotional category as that of the starting frame. The results showed a robust overshoot response bias: the neutral expression in the last frame of the joy-to-‘neutral’ videos was judged as slightly angry, while the identical neutral expression in the last frame of the Anger-to-neutral videos was judged as slightly happy, confirming the results by Jellema et al. (2011). This finding was quite consistent across actors. Importantly, the conditions in which the endpoint was not neutral, but 10% joy or 10% anger, demonstrated that this response bias was strong enough to overcome a 20% difference in emotional intensity, as the 10% joy endpoint in the joy-videos was rated as significantly more angry than the 10% anger endpoint in the anger-videos.

In the following five experiments the mechanism underpinning this perceptual distortion of facial expressions induced by the perceptual history was investigated, by attempting to disentangle the contributions of bottom-up visual processes from top-down emotional anticipation.

2.3 Experiment 2

In Experiment 2 the identity of the agent was changed in the last frame of the video-clips. Therefore the judgment of the final expression was made from an identity that was new to the observer; i.e. someone for whom no immediate perceptual history was available. The rationale was that if the overshoot bias would result from sequential contrast/context effects then a change in identity should not affect the bias, as the contrast remained intact. Studies of contrast effects showed that it does not matter whether the emotion depicted in the anchor and target stimuli are of the same or different identities (e.g. Tanaka-Matsumi et al., 1995; Thayer, 1980). Similarly, if the overshoot bias would result from extrapolation (RM) on an underlying positive-negative valence dimension, then it should also not be affected by the change in identity as the underlying valence dimension remains intact. If the bias was due to adaptation to the intense facial expression in the first frame (300 ms), resulting in an aftereffect in opposite direction, then a change in identity should again not affect the bias, as the facial expressions aftereffects typically generalise over different identities, although to a less extent (e.g. Fox & Barton, 2007; Ellamil, Susskind & Anderson, 2008). Thus, if the response bias would be reduced or removed due to the identity change, than that would suggest that these low-level mechanisms were not the only elements causing the response bias. Such a result (bias removal) would, however, be consistent with an emotional anticipation account. Emotional anticipation would predict the absence of an overshoot bias as the observer cannot anticipate the emotional state of someone on the basis of a perceptual history that pertains to a different identity.

Even though care was taken to limit the perceptual shift due to the identity transition, a certain amount of shift was unavoidable. In theory, this might have disrupted the perceptual flow causing the perceptual system to 'reset' (cf. Enns et al., 2010). Therefore, additionally, a version of Experiment 2 was conducted in which the identity change was smooth and gradual (smooth-identity-change experiment).

2.3.1 Method

2.3.1.1 Participants

The two versions of Experiment 2 were performed on two different samples of participants. All participants were undergraduate Psychology students at Hull University. After applying exclusion criteria (see below), twenty-three participants were included in the instant-identity-change experiment (age, M = 19.3 years, SD = 1.53 years; 19 females, 4 males), and twenty participants in the smooth-identity-change experiment (age, M = 19.4 years, SD = 2.56 years, 17 females, 3 males). None of the participants took part in any of the other experiments. All had normal or corrected-to-normal vision, gave informed consent, and received course credit for taking part.

2.3.1.2 Stimuli

First, in the same identity condition, one identity was shown throughout the video, from 100% joy or anger to neutral (Figure 2.4, top panel). This condition was identical to the condition shown in Experiment 1 where the endpoint was 0%.

In the identity-change conditions, videos started with an actor (identity-A) displaying a 100% emotional expression (joy or anger), which ended with the neutral expression depicted by a different actor (identity-B; Figure 2.4, middle and bottom panels).

In the instant identity-change experiment, the identity changed from the last but one to the last frame (Figure 2.4, middle panel). The entire duration was 870 ms (first frame for 300 ms, 9 interpolated frames for 30 ms each, and last frame for 300 ms). The face images of identities A and B were resized to match in outer dimensions as much as possible, with the eyes presented at a fixed location on the screen, to try and limit the extent of geometrical shift between identities.

In the smooth identity-change experiment, the identity transition was morphed by creating three interpolations between identity-A at 10% happy or angry and identity-B at 0%, and inserting these in the sequence directly before the last, neutral, frame (Figure 2.4, bottom panel). The three inserted frames consisted of 80%A-20%B, 50%A-50%B and 20%A-80%B. Each lasted 20 ms, increasing the total duration of the sequence to 930 ms. Each actor was combined with four other actors, two of which were of the same and two of opposite gender.

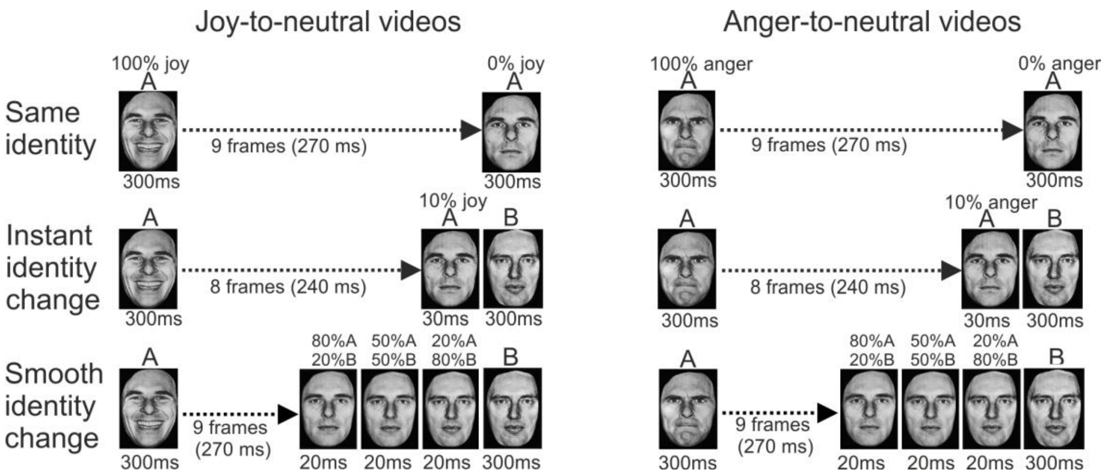

Figure 2.4. Schematic illustration of the stimulus presentations in Experiment 2. Examples of the same identity condition (top), the instant identity-change condition (middle) and the smooth identity-change condition (bottom), for Joy-to-neutral (left side) and Anger-to-neutral (right side) perceptual histories.

2.3.1.3 Experimental procedure

Both the instant and smooth identity-change experiments started with a calibration phase, in which the participants rated the static neutral expression of each actor presented for 600 ms, using the same 5-point scale (ranging from slightly angry, via neutral, to slightly happy), similarly to Experiment 1. Then, following 8 practice trials, 64 experimental (identity-change) and 64 control (same identity) trials were presented in random order (8 actors x 2 perceptual histories x 4 repetitions). The participants' task was to evaluate the last frame of the video-clips on the same 5-point scale as used in the calibration phase. Further procedures and apparatus were as in Experiment 1.

2.3.1.4 Data reduction and analysis

Exclusion criteria were the same as used in Experiment 1. Trials in which the RTs fell below 250 ms or above 3000 ms were removed (4.9% in the instant-change and 6.9% in the smooth-change experiment).

Participants were excluded if more than 25% of the RTs values were outside this range (n = 2, in both experiments). The $\pm$2.5 SD rule to the mean difference ratings excluded another three participants (one in the instant-change and two in the smooth-change experiment).

As in Experiment 1, the calibration scores were used to adjust the scores in the experimental trials and all statistical analyses were performed on the calibrated scores.

2.3.2 Results

2.3.2.1 Instant change in identity

A 2x2 repeated measures ANOVA with Identity-change (no change vs. change) and Perceptual history (Joy-to-neutral vs. Anger-to-neutral) and as the within-subject factors showed significant main effects of Identity change ($F(1, 22) = 7.9$, $p = .010$, $\eta_p^2 = .264$) and Perceptual history ($F(1, 22) = 41.2$, $p < .0001$, $\eta_p^2 = .652$).

Importantly, the interaction factor was highly significant ($F(1, 22) = 18.8$, $p < .0001$, $\eta_p^2 = .461$; Figure 2.5A). Similarly to Experiment 1, in the same identity condition, neutral faces were perceived as more angry in the Joy-to-neutral videos (M = 2.82, SD = 0.23) than in the Anger-to-neutral videos (M = 3.32, SD = .26), ($t(22) = -7.13$, $p < .0001$), and each differed significantly from 3.00 (joy: $t(22) = -3.69$, $p = .001$; anger: $t(22) = 5.87$, $p < .0001$). However, in the identity-change condition the ratings for the neutral expressions in the two perceptual histories did not differ ($t(22) = -1.15$, $p = .263$). Thus, a change of identity of the actor in the last frame of the video-sequence effectively removed the overshoot bias. The consistency of the lack of a response bias across actors in the identity-change condition is illustrated in Figure 2.5 B.

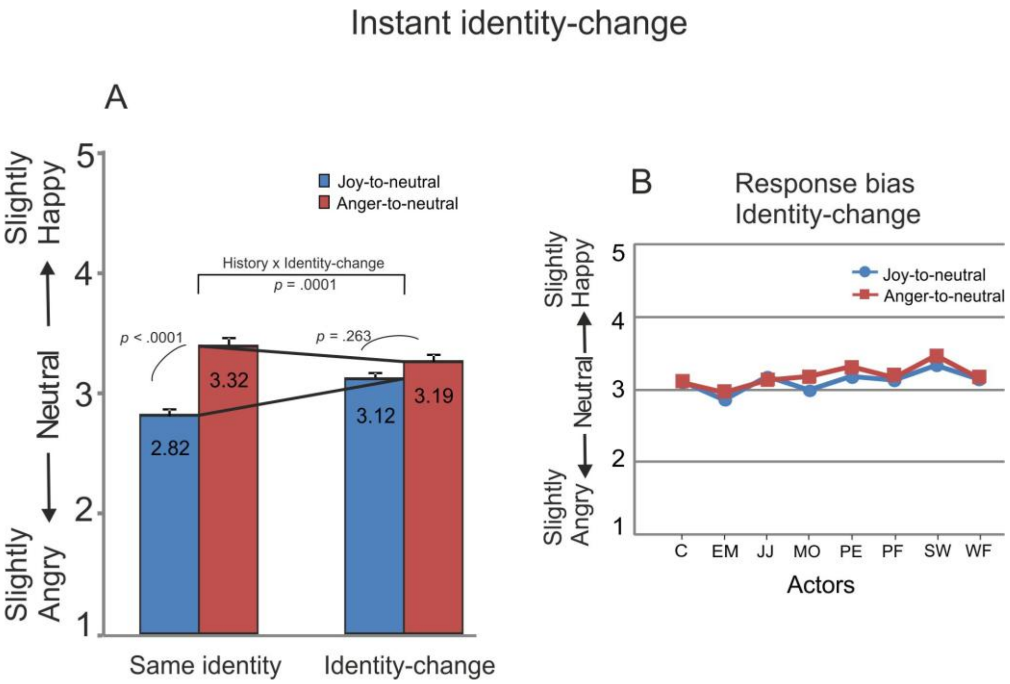

Figure 2.5. Results of the Instant identity-change version of Experiment 2. (A) Scores for the neutral expressions at the end of the Joy-to-neutral and Anger-to-neutral sequences in the same identity (left) and instant identity-change conditions (right). Error bars indicate SEM. (B) Ratings for each of the eight actors in the instant identity-change condition are shown to illustrate the consistency in the lack of a response bias across actors.

2.3.2.2 Smooth change in identity

A 2x2 ANOVA, similar to the one performed for the Instant identity-change experiment (section 2.3.2.1), showed no significant main effect for the Identity-change factor ($F(1, 19) = .25$, $p = .62$, $\eta_p^2 = .013$) but a robustly significant main effect for Perceptual history ($F(1, 19) = 75.3$, $p < .0001$, $\eta_p^2 = .79$). The interaction factor was highly significant ($F(1, 19) = 33.9$, $p < .0001$, $\eta_p^2 = .64$; Figure 2.6A). In the same identity condition the perceptual histories again differed significantly from each other ($t(20) = -8.68$, $p < .0001$) and from 3.00 (*p's* < .01). In the smooth identity-change condition, even though the mean ratings for the neutral expression in the joy and anger videos overlapped considerably (Figure 2.6B), the ratings in two perceptual histories did differ significantly from each other $t(19) = -4.21$, $p < .001$). However, neither perceptual history differed from 3.00 (Joy, $t(19) = -1.5$, $p = .15$; Anger, $t(19) = .72$, $p = .48$).

Importantly, doing a comparison between the smooth identity change and the abrupt identity change versions, a 2x2x2 mixed ANOVA with Perceptual History (Joy-to-neutral vs. Anger-to-neutral) and Identity-change (no change vs. change) as the within subjects factors and Group (smooth vs. abrupt) as the between subjects factor revealed a main effect of Perceptual history ($F(1, 41) = 100.89$, $p < .0001$, $\eta_p^2 = .711$) and a main effect of Group ($F(1, 41) = 4.50$, $p < .05$, $\eta_p^2 = .099$), whereas the main effect of Identity-change was nearly significant ($F(1, 41) = 3.72$, $p < .061$, $\eta_p^2 = .083$). Remarkably, the 2-way interaction Perceptual history by Condition was significant ($F(1, 41) = 41.17$, $p < .0001$, $\eta_p^2 = .018$), whereas the 2-way interactions with Group were non-significant (all *p*s > .3). The 3-way interaction Perceptual history by Identity by Group was also not significant ($F(1, 41) = .75$, $p = .392$, $\eta_p^2 = .018$). This means that the amount of overshoot as found in the smooth version is really subtle.

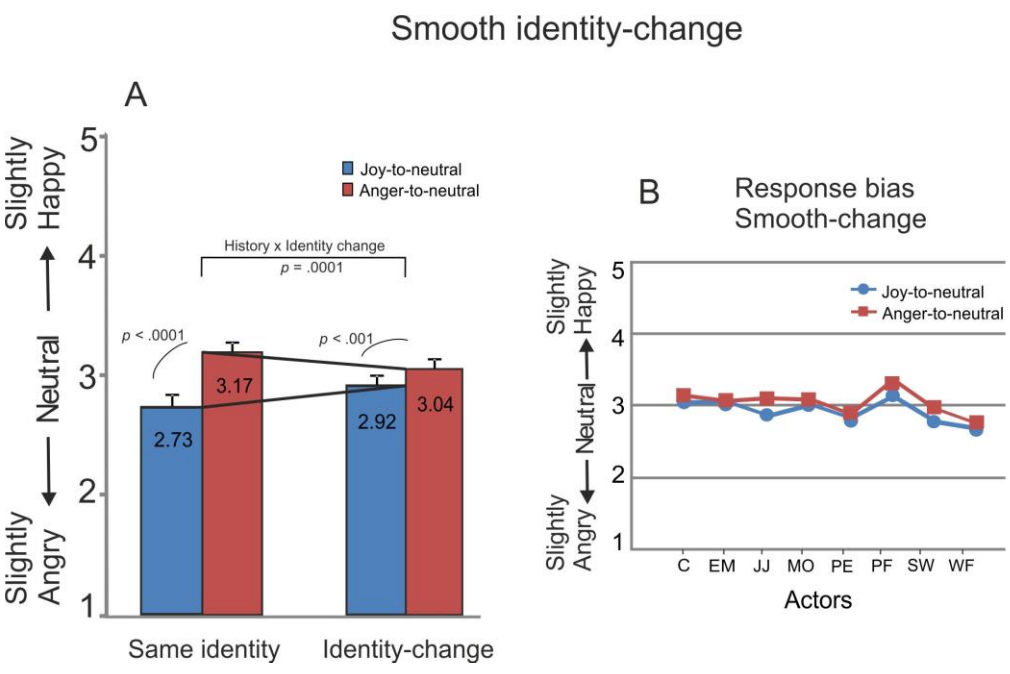

Figure 2.6. Results of the Smooth identity-change version of Experiment 2. (A) Scores for the neutral expressions at the end of the Joy-to-neutral and Anger-to-neutral sequences in the same identity (left) and smooth identity-change (right) conditions. Error bars indicate SEM. (B) Ratings for each of the eight actors in the smooth identity-change condition are shown to illustrate the consistency in (lack of) response biases across actors.

2.3.3 Discussion

In Experiment 2 a change in the actor's identity was introduced at the end of the video-clips to test whether the overshoot response bias could have been generated by sequential contrast/context effects, by extrapolation (RM) on an underlying positive-to-negative valence dimension, or by adaptation to the expression displayed in the first frame.

The current results showed that for an overshoot response bias to occur, the identity of the actor should remain the same during the perceptual history: a change in identity in the last frame of the sequence effectively removed the bias. The removal of the bias was more evident when the identity change was instantaneous, while some bias seeped through when the identity-change was smoothed. The most likely explanation for this seeping through is that the participants in some cases did not notice the change in identity due to the morphing procedure and believed actors A and B to be one and the same person. In contrast, in the instant condition the identity change was always noticed. This was confirmed by debriefing of participants.

The results suggest that the proposed low-level visual mechanisms do not significantly contribute to the generation of the response bias in our sample of typically-developed individuals. Influences of contrast effects to the evaluation of facial expressions are known to be independent on the identity of the actors (Thayer, 1980). Therefore, the significant reduction (Smooth identity-change experiment) and removal (Instant identity-change experiment) of the bias suggests that the large bias in the same-identity condition was not caused by sequential contrast effects. If it was caused by contrast effects, there should have been an overshoot bias in the identity-change conditions as well. The contrasts in facial expression and emotional valence between the 100% emotional expression and the neutral expression were equally strong in same identity and identity-change conditions.

The results further suggest that extrapolation on an underlying positive-negative valence dimension (RM) did not cause the overshoot bias in the same identity condition either, as an identity change does not change the underlying valence dimension.

It seems the results also ruled out adaptation as a major cause of the bias as the change in identity should not dramatically affect the aftereffects induced by adaption to a facial expression. Although some decrease in aftereffect may be anticipated (Fox & Barton, 2007; Ellamil et al., 2008), a significant adaptation effect should still be present. However, the instant identity-change experiment revealed no overshoot bias whatsoever. Moreover, the perceptual histories used in our paradigm were fairly short (the initial frame depicting the maximally intense emotion was presented for just 300 ms). Several studies (Leopold et al., 2001; Rhodes et al., 2003; Webster et al., 2004) reported that face afteraffcts only appear with adaptor durations ranging from several seconds to minutes, which renders adaptation a less likely candidate. For example, Leopold et al. (2001) reported that periods shorter than 5 s failed to produce facial aftereffects. One exception is provided by Rijsbergen, Jannati, & Treves (2008) who found aftereffects for emotional expression with adaptor primes of 24 ms. However, it remains unclear whether the perceptual distorsion found by Rijsbergen et al. (2008) may rather be due to contrast-effects.

The results are compatible with the idea that the overshoot bias was due to top-down emotional anticipation. Actor B in the last frame was someone for whom no perceptual history had been witnessed, so the observer did not know anything about B’s emotional state other than that B had a neutral expression, and therefore rated B as neutral.

In Experiment 3 the role played by contrast/context effects in bringing about the overshoot bias was further investigated.

2.4 Experiment 3

Experiment 3 was designed to further assess a possible contribution of contrast effects to the overshoot response bias but by using a different rationale than in Experiment 2. The video-sequence was modified so that it started from a neutral expression and morphed via an expression of maximal joy or anger back to the same neutral expression (Loop condition). If the response bias found in the previous experiments was due to the contrast between the to-be-evaluated expression in the last frame and the ‘anchor’ stimulus in the first frame presented for 300 ms, then in the Loop condition one would expect the last neutral expression to be judged as neutral and not obtain an overshoot bias, despite the second half of the sequence (from maximally intense emotion to neutral) being identical to the sequence in the previous experiments.

2.4.1 Method

2.4.1.1 Participants

Twenty-eight undergraduate Psychology students at Hull University took part in the experiment. After applying exclusion criteria, twenty-four participants were included in the analysis (age, M = 19.92 years, SD = 2.04; 15 females, 9 males). None of the participants took part in any of the other experiments. All participants had normal or corrected-to-normal vision, gave informed consent, and received course credit for taking part.

2.4.1.2 Stimuli

In the No loop condition, the video-clips were identical to those used for the same-identity condition of Experiment 2 (Joy/Anger-to-neutral). In the Loop condition, a modification was implemented such that the video-clips started with a neutral expression, and morphed via maximal joy (or anger) back to the same neutral expression. In this condition, the morphing sequence consisted of nineteen interpolated frames, each 30 ms long. The first and the last frames both lasted 300 ms, making the entire sequence last for 1170 ms (see Figure 2.7).

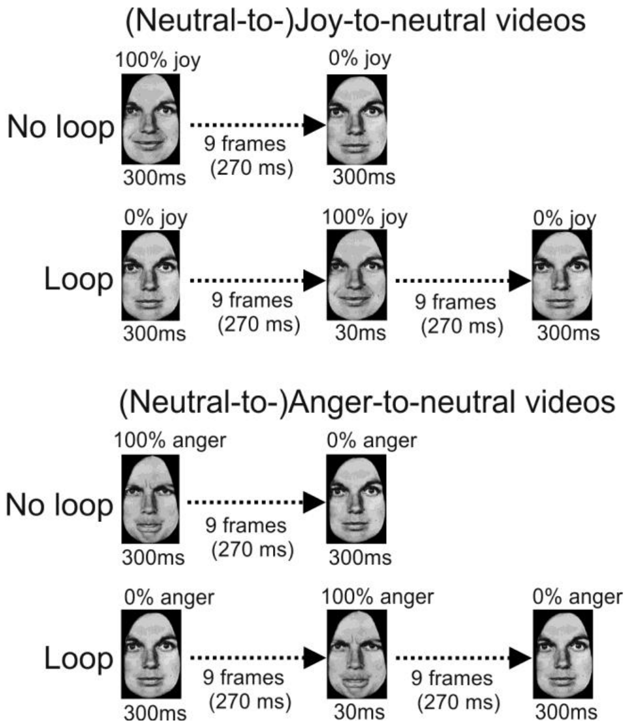

Figure 2.7. Schematic illustration of the stimulus presentations in Experiment 3. (Top panel) Joy-to-neutral sequence in the No loop condition (top row) and Neutral-to-joy-to-neutral sequence in the loop condition (bottom row). (Bottom panel) Similar sequences but for Anger-to-neutral (top row) and Neutral-to-anger-to-neutral (bottom row).

2.4.1.3 Experimental procedure

Procedures were as described in Experiment 1 and 2. Participants first rated the static neutral expression of each actor presented for 600 ms by using the 5-point scale.

Then, following 8 practice trials, one experimental block was presented with 96 randomized trials: 8 actors x 2 perceptual histories x 2 conditions x 3 repetitions. The participants' task was to evaluate the last frame of the video-clips on the same 5-point scale as used in the calibration. Further procedures and apparatus were as in the Experiments 1 and 2.

2.4.1.4 Data reduction and analysis

Exclusion criteria were as in the previous experiments. Trials in which RTs fell below 250 ms or above 3000 ms were removed (7.9%). Participants were excluded if more than 25% of the RTs values fell out of the mentioned range (n = 3) and when the mean rating fell out of a $\pm$2.5 SD range (n = 1). The analyses were performed on the calibrated scores as in the previous experiments.

2.4.2 Results

The 2x2 repeated measures ANOVA with Condition (No loop vs. Loop) and Perceptual history (Joy-to-neutral videos vs. Anger-to-neutral videos) as the within-subjects factors showed no significant main effect for Condition ($F(1, 23) = .35$, $p = .56$, $\eta_p^2 = .015$) and Perceptual history ($F(1, 23) = .024$, $p = .88$, $\eta_p^2 = .001$), but a highly significant Condition by Perceptual history interaction ($F(1, 23) = 21.4$, $p < .0001$, $\eta_p^2 = .48$; Figure 2.8). In the No loop condition the typical overshoot effect was found; the mean ratings of the neutral faces was 2.79 in the Joy-to-neutral and 3.20 in the Anger-to-neutral sequences, which differed significantly from each other ($t(23) = 4.18$, $p < .0001$). However, in the Loop condition the reversed pattern emerged: the neutral expression at the end of the Neutral-to-joy-to-neutral video was evaluated as slightly happy (M = 3.24, SD = 0.53), and at the end of the Neutral-to-anger-to-neutral videos as slightly angry (M = 2.80, SD = 0.47). These evaluations differed significantly from each other ($t(23) = 2.60$, $p = .016$) and each differed significantly from 3.00 (neutral-joy-neutral: ($t(23) = 2.50$, $p = .020$); neutral-anger-neutral ($t(23) = -2.16$, $p = .041$; two tailed). The response bias, which was in opposite directions in the No loop and the Loop conditions, was consistent across actors as illustrated in Figure 2.8. The reversed response bias in the Loop condition might be called an “undershoot” response bias.

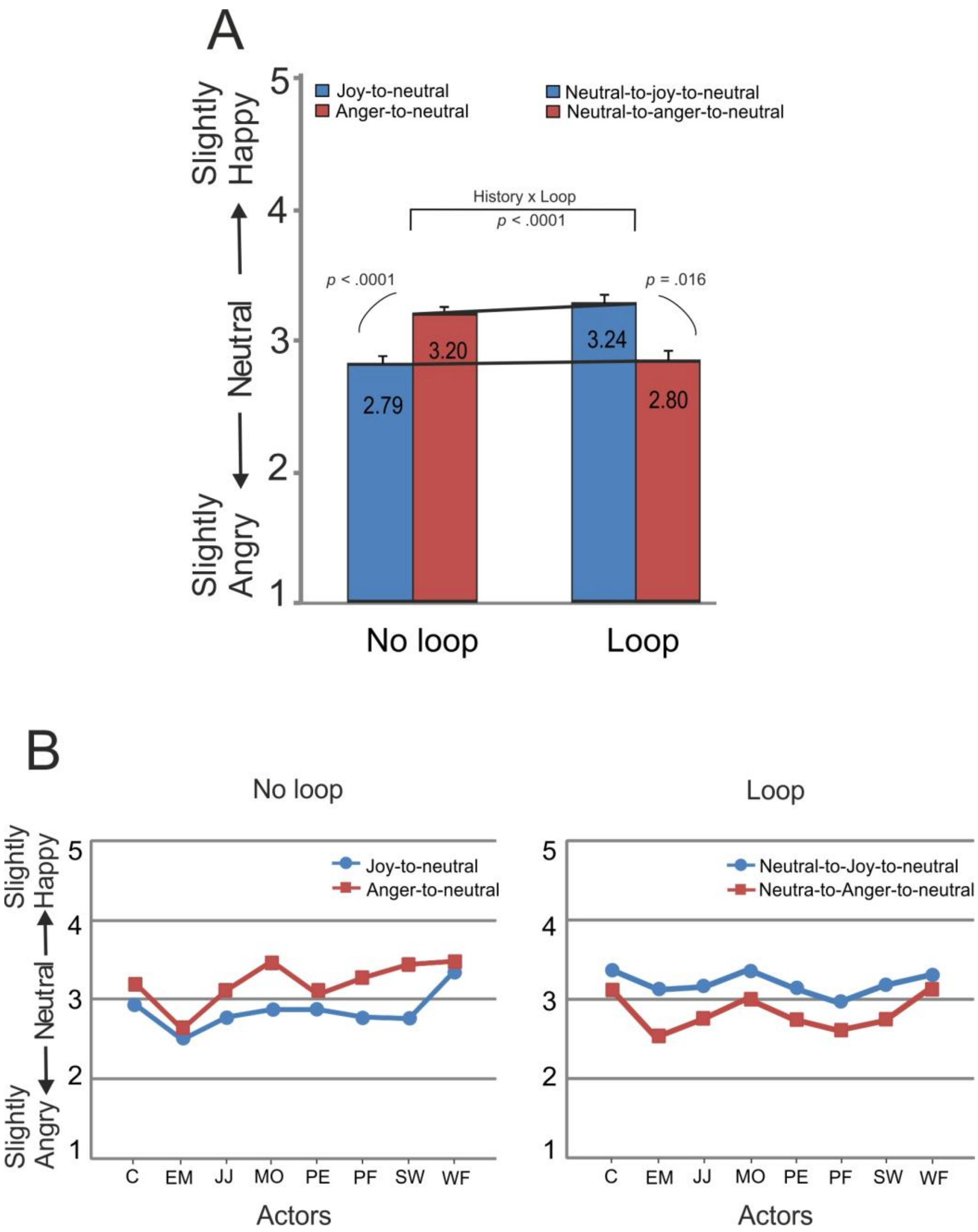

Figure 2.8. Results of Experiment 3. (A) Ratings of the neutral expressions at the end of the video-sequences in the No loop and in the Loop conditions. Error bars indicate SEM. (B) Ratings for each of the eight actors in the No loop and Loop conditions are shown to illustrate the consistency across actors.

2.4.3 Discussion

The experimental manipulation in Experiment 3 consisted of removal of the 300 ms frame depicting the maximally intense emotion at the start of the sequences and replacing it by a 300 ms frame depicting a neutral expression, which gradually morphed into a maximally intense emotion, and then morphed back into the same neutral expression as where it started from. Its aim was to further test for a possible contribution of sequential contrast effects to the overshoot response bias. The results of this experiment established that the contrast between the initial frame (300 ms duration) and the last, to-be-evaluated, frame (300 ms) of the sequence did not play a significant role in bringing about the overshoot response bias. In the Loop condition there was no contrast between the initial and last frames (both showed identical neutral expressions) and therefore the sequential contrast hypothesis predicted the absence of a response bias. However, a response bias was obtained but in opposite in direction to the one in the No loop condition (i.e. an undershoot response bias).

The findings of Experiment 3 are also not compatible with an extrapolation on an underlying positive-negative valence dimension (or RM) being the cause for the overshoot response bias, as this would have predicted an overshoot bias for the Loop condition (i.e. a continuation of the second half of the video).

In principle, emotional anticipation, i.e. the tendency to keep track of the other's emotional state of mind and to predict what the emotional state is most likely to be in the immediate future, is able to explain the reversed bias direction (undershoot). The perceptual histories in the loop condition, differently from the control condition, involve whole expressions (onset and offset of a smile or frown). In our daily life we rarely attribute a negative state of mind to a person who smiles at us while it is difficult to attribute a positive state of mind to a person who frowns. Thus, it could be that observers perceived the brief smile as a positive social sign, indicating the agent possessed a positive attitude towards them, and consequently evaluated the agent positively, as we tend to like people who like us (cf. Jones, DeBruine, Little, Conway, & Feinburg, 2006). Similarly, the brief frown may have constituted a negative social sign resulting in the notion that the agent held a negative disposition toward the observer, and hence the negative evaluation of the neutral expression.

2.5 Experiment 4

Experiment 4 was specifically designed to examine the possible contribution of RM to the overshoot response bias. Yoshikawa and Sato (2008) presented short video-clips of neutral (0%) expressions that morphed into more intense expressions (70%, 80% or 90%). Video-sequences of 21 frames at equal steps of 4% intensity change were administered with four different velocity conditions: 210 ms (10 ms/frame), 420 ms (20 ms/frame), 840 ms (40 ms/frame) and 1680 ms (80 ms/frame).

The authors reported that participants overestimated the expression of the last frame of the sequence and that the size of this effect was positively correlated with the speed of the intensity change. These results were interpreted in favour of an RM effect, which is the phenomenon that an observer’s memory for the final position of a moving target is displaced further along the observed trajectory (Freyd, 1987; Freyd & Finke, 1984).

Similarly, in the ‘overshoot’ paradigm with Joy-to-neutral and Anger-to-neutral videos, one could expect an RM effect operating on the dynamic facial features. The motion contained in the Joy-to-neutral videos could have caused a displacement of the memory of the last neutral expression further along its trajectory, crossing the category boundary. Similarly, the motion contained in the Anger-to-neutral videos could have displaced the memory of the same last neutral expression beyond its category boundary, resulting in the recollection of a slightly happy expression. In Experiment 4 the speed of the video-clips was manipulated to test whether this would modulate the overshoot effect. If the amount of the overshoot bias does not positively correlate with the speed of the video-clips then that would make it unlikely that the overshoot bias is related to RM.

2.5.1 Method

2.5.1.1 Participants

Twenty-two undergraduate Psychology students at Hull University took part in the experiment. After applying exclusion criteria, twenty participants were included in the analysis (age, M = 21.3 years, SD = 1.77; 13 females, 7 males). None of the participants took part in any of the other experiments. All had normal or corrected-to-normal vision, gave informed consent, and received course credit for taking part.

2.5.1.2 Stimuli

Joy-to-neutral and Anger-to-neutral video-clips of the same eight actors were presented in four different velocity conditions, which were called: fast, normal, slow and very slow. The fast condition consisted of only one interpolated frame of 50% intensity between the first maximally happy or angry (100%) frame and the last neutral (0%) frame of the sequence. The interpolated frame lasted 15 ms, while the first and the last frames both lasted 300 ms (total duration of the entire sequence:

615 ms). The normal condition contained nineteen interpolated images in between the full-blown expression of joy or anger (100%) and the neutral expression (0%), at equal steps of 5% intensity change.

The 19 interpolated frames lasted 15 ms each one (total duration: 285 ms).

The velocity of the normal condition was fairly similar to the velocity of the videos used for the control conditions in the previous experiments (30 ms by 9 frames = 270 ms). The first (100%) and the last (0%) frames both lasted 300 ms (total duration of the entire sequence: 885 ms). In the slow condition the same nineteen interpolated images lasted 30 ms each one (total duration: 570 ms) and the first 100% happy or angry frame and the last 0% neutral frame both lasted 300 ms (total duration of the entire sequence: 1170 ms). Finally, in the very slow condition the duration of the nineteen interpolated frames was increased to 45 ms each (total duration 855 ms), making the total duration of the entire video, including the first 100 % and the last 0% frames, 1455 ms. The four conditions are illustrated in Figure 2.9.

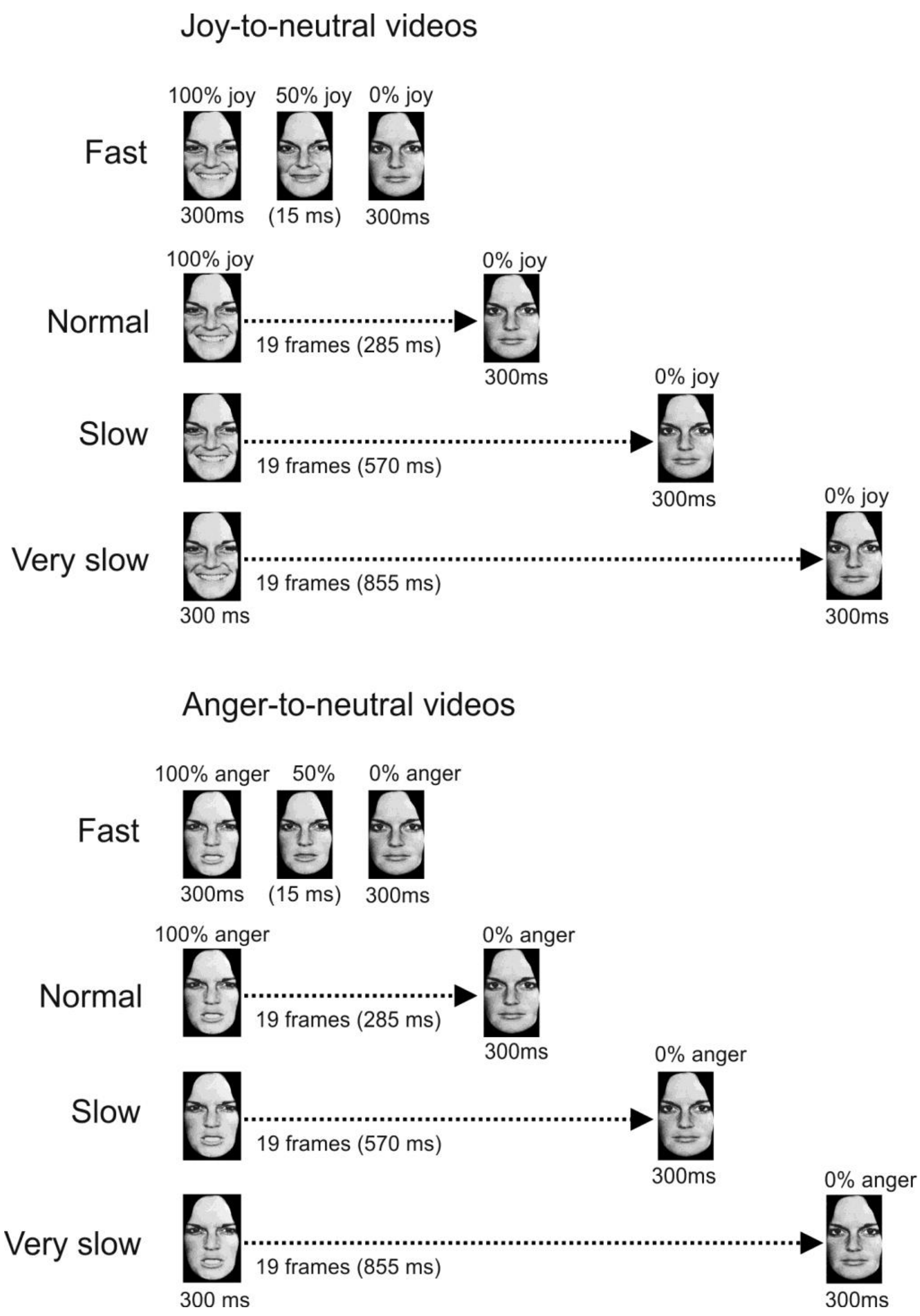

Figure 2.9. Schematic illustration of the stimulus presentations in Experiment 4. Joy-to-neutral videos (top panel) and Anger-to-neutral videos (bottom panel) in the four velocity conditions (fast, normal, slow and very slow).

2.5.1.3 Experimental procedure

Procedures were the same as in the previous experiments, except that Experiment 4 consisted of 128 randomized trials (8 actors x 2 perceptual histories x 4 speeds x 2 repetitions).

2.5.1.4 Data reduction and analysis

Exclusion criteria were as in the previous experiments. Trials in which RTs fell below 250 ms or above 3000 ms were removed (4.5%). Participants were excluded if more than 25% of the RTs values fell out of the mentioned range (n = 0) and when the mean rating fell outside a $\pm$2.5 SD range (n = 2). As in the previous experiments all the analyses were performed on the calibrated scores.

2.5.2 Results

The results are shown in Figure 2.10. The 2x4 repeated measures ANOVA with Perceptual history (Joy-to-neutral videos vs. Anger-to-neutral videos) and Velocity (fast vs. normal vs. slow vs. very slow) as the within-subjects factors showed a strong significant effect for Perceptual history ($F(1, 19) = 35.01$, $p < .0001$, $\eta_p^2 = .648$), but no significant main effect for Velocity ($F(3, 57) = 1.01$, $p = .395$, $\eta_p^2 = .050$). Importantly the Velocity by Perceptual history interaction was not significant ($F(3, 57) = .795$, $p = .502$, $\eta_p^2 = .040$). Paired sample t-tests revealed a robust overshoot bias in each of the 4 velocity conditions (Fast: $t(19) = -4.44$, $p < .0001$; Normal): $t(19) = -4.04$, $p = .001$; Slow: $t(19) = -6.45$, $p < .0001$; and Very slow : $t(19) = -5.15$, $p < .0001$; $\alpha = 0.012$). The mean scores in each velocity condition for each actor separately are also illustrated in Figure 2.10.

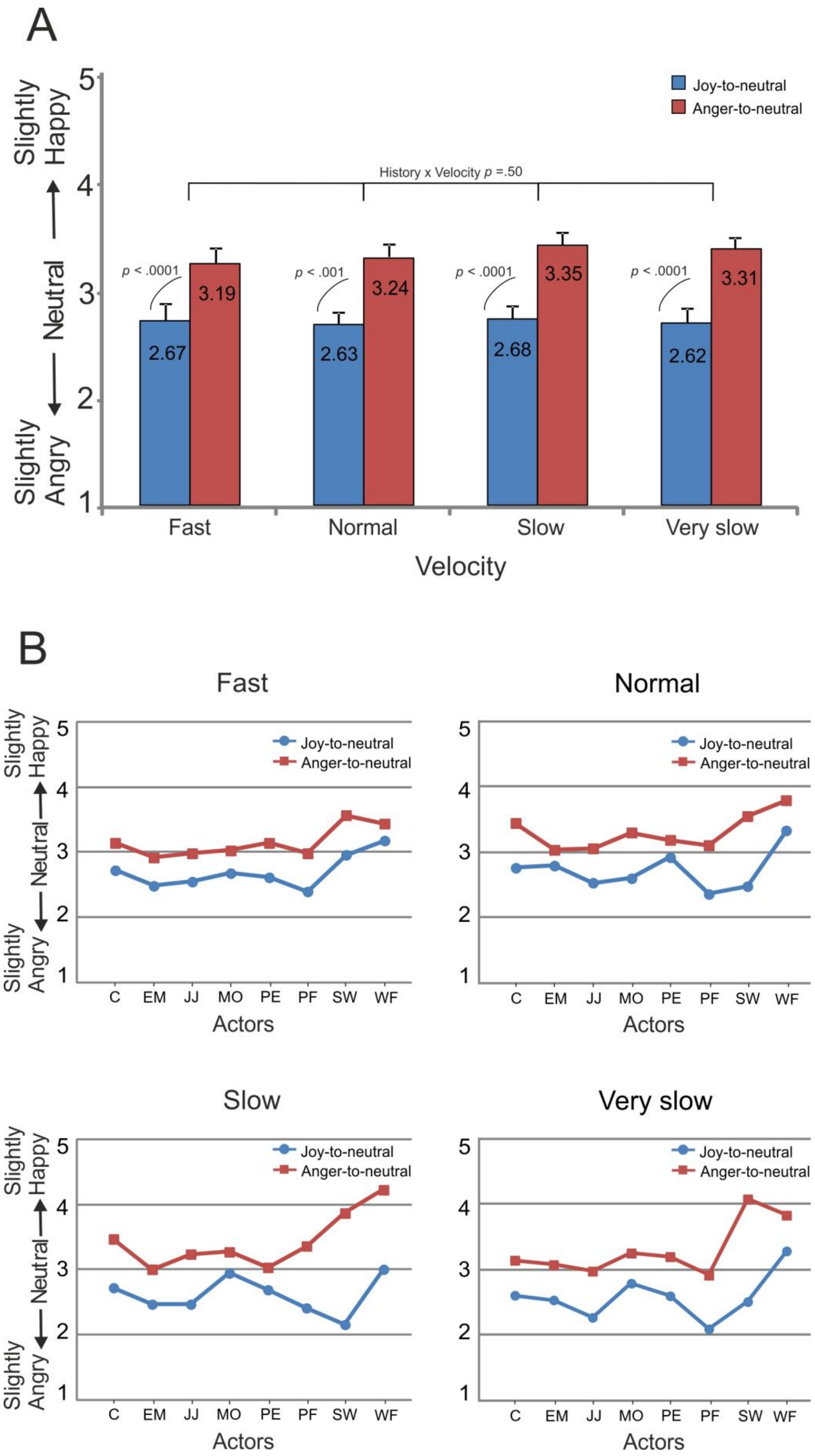
A
Joy-to-neutral
Anger-to-neutral
History x Velocity p =.50
p < .0001
p < .001
p < .0001
p < .0001
2.67
3.19
2.63
3.24
2.68
3.35
2.62
3.31
Slightly Happy
Neutral
Slightly Angry
Fast
Normal
Slow
Very slow
Velocity
B
Fast
Normal
Slow
Very slow
C EM JJ MO PE PF SW WF
Actors

Figure 2.10. Results of Experiment 4. (A) Ratings of the neutral expressions at the end of the video-sequences at fast, normal, slow and very slow velocities. Error bars indicate SEM. (B) Ratings for each of the eight actors in the four conditions (fast, normal, slow and very slow) are shown to illustrate the consistency across actors.

2.5.3 Discussion

Experiment 4 was conducted to test the possibility that the overshoot response bias found at normal velocity (see the previous experiments) could be explained in terms of RM rather than in terms of an emotional mindreading/anticipation mechanism. Sato and Yoshikawa (2008) used video-sequences, which morphed from neutral to an intense expression, with four different velocity conditions: 210 ms, 420 ms, 840 ms and 1680 ms. In their study participants overestimated the intensity of the expressions depicted in the last frame when the video-clips were fast (210 ms and 420 ms), while this effect was not found with slower videos (840 ms and 1680 ms). Experiment 4 presented here showed that in our paradigm the overshoot effect is not modulated by the speed of the video-clips. The four velocity conditions adopted in the current experiment were fairly comparable to the ones adopted by Yoshikawa and Sato (2008). Therefore, it is unlikely that the response bias found in the current study can be explained in terms of RM.

Moreover, as noted in the Methods section of Experiment 1 (2.2.1.3), the RM paradigm applied by Yoshikawa and Sato (2008) presents some methodological differences with the ‘overshoot’ paradigm. First of all, in the study by Yoshikawa and Sato participants were instructed to select, using an interactive slider, the emotional intensity out of a sequence of different possible emotional intensities (i.e. ranges of

20%-120%, 30%-130%, 40%-140%), which they believed corresponded best to the expression they saw at the end of the video-clip. This procedure could itself have affected participants' choices as this selection procedure created new perceptual histories, affecting the perception of the test intensities themselves. In the 'overshoot' paradigm, participants needed to make perceptual evaluations using a verbal 5-point scale, of the last (neutral) frame of the sequence, without being showed any other expressions for comparison. This removed the possibility of 'contamination' by the test stimuli. Further, in contrast to the reversed video-sequences (i.e. neutral-to-joy) used by Yoshikawa and Sato (2008), the overshoot effect obtained with Joy-to-neutral and Anger-to-neutral videos entailed a crossing of the category boundary.

2.6 Experiment 5

In the video-sequences presented so far, many geometrical features in the face changed in a continuous and smooth manner. To examine to what extent the overshoot response depended on the continuous, uninterrupted flow of motion up to the last, neutral frame that had to be evaluated, in Experiment 5 the video-sequence was disrupted by inserting a mask (consisted of a 400 ms long frame) immediately before this last frame. If the overshoot bias is determined by an extrapolation of the facial geometries then interruption of the video flow would remove or reduce the bias. For example, the U-shaped mouth in the smile gradually transformed into a flat-shaped mouth in the neutral expression that might be extrapolated into a slightly inverted U shape, which would give the face a slightly angry look. However, insertion of the mask would not affect the anticipation of the emotional state of the actor, and therefore the emotional anticipation explanation would predict to find an overshoot bias.

2.6.1 Method

2.6.1.1 Participants

Sixteen undergraduate Psychology students at Hull University took part. After applying exclusion criteria (see below), fifteen participants were included in the analysis (age, M = 20.2 years, SD = 3.0; 12 females, 3 males). None of the participants took part in any of the other experiments. All had normal or corrected-to-normal vision, gave informed consent, and received course credit for taking part.

2.6.1.2 Stimuli

The same Joy-to-neutral and Anger-to-neutral video-clips were presented as in the normal velocity condition in Experiment 4, which also corresponds to the control condition of the previous experiments. A 400 ms long mask, consisting of an oval grid in grey scale, was inserted in the sequence directly before the last frame (Figure 2.11). The contour of the oval shape matched that of the faces. Due to insertion of the mask the total duration of the sequence was 1270 ms; all other stimulus parameters were as in the previous experiments. The control condition was identical to the mask condition except that no mask was inserted (Figure 2.11).

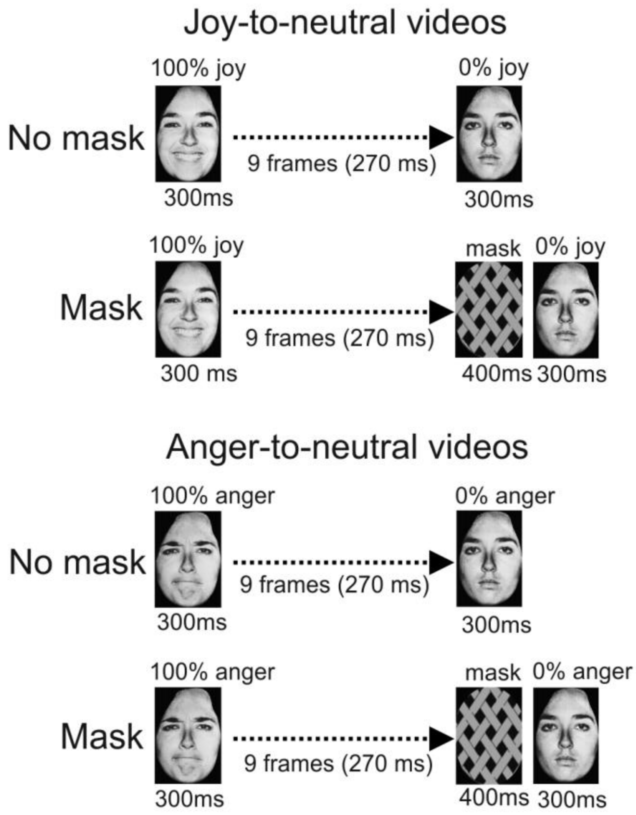

Figure 2.11. Schematic illustration of the stimulus presentations in Experiment 5. Joy-to-neutral videos (top panel) and Anger-to-neutral videos (bottom panel) in the No mask and in the Mask conditions are shown.

2.6.1.3 Experimental procedure

As for the previous experiments, participants started with a calibration phase, rating the neutral expression of each actor presented for 600 ms using the 5-point scale. Following 8 practice trials, one experimental block was presented with 64 trials, 32 for the no mask condition and 32 for the mask condition (8 actors x 2 perceptual histories x 2 repetitions) in random order. The participants' task was to evaluate the last frame of the video-clips on the same 5-point scale used in the calibration. Further procedures and apparatus were as in the previous experiments.

2.6.1.4 Data reduction and analysis

The same criteria as for the previous experiments were adopted. Trials in which the RTs fell below 250 ms or above 3000 ms were removed (8%). Participants were excluded if more than 25% of the RTs values were out of the mentioned range (n = 0). The $\pm$2.5 SD rule to the mean difference ratings excluded one participant. The analyses were performed on the calibrated scores as in the previous experiments.

2.6.2 Results

A 2X2 repeated measures ANOVA with Condition (No mask vs. Mask) and Perceptual history (Joy-to-neutral vs. Anger-to-neutral) as factors showed a robustly significant main effect of Perceptual history ($F(1, 14) = 47.50$, $p < .0001$, $\eta_p^2 = .77$). The factor Condition was not significant ($F(1, 14) = 1.86$, $p = .194$, $\eta_p^2 = .117$), and the Condition by Perceptual history interaction ($F(1, 14) = 1.68$, $p = .215$, $\eta_p^2 = .107$) was not significant (Figure 2.12).

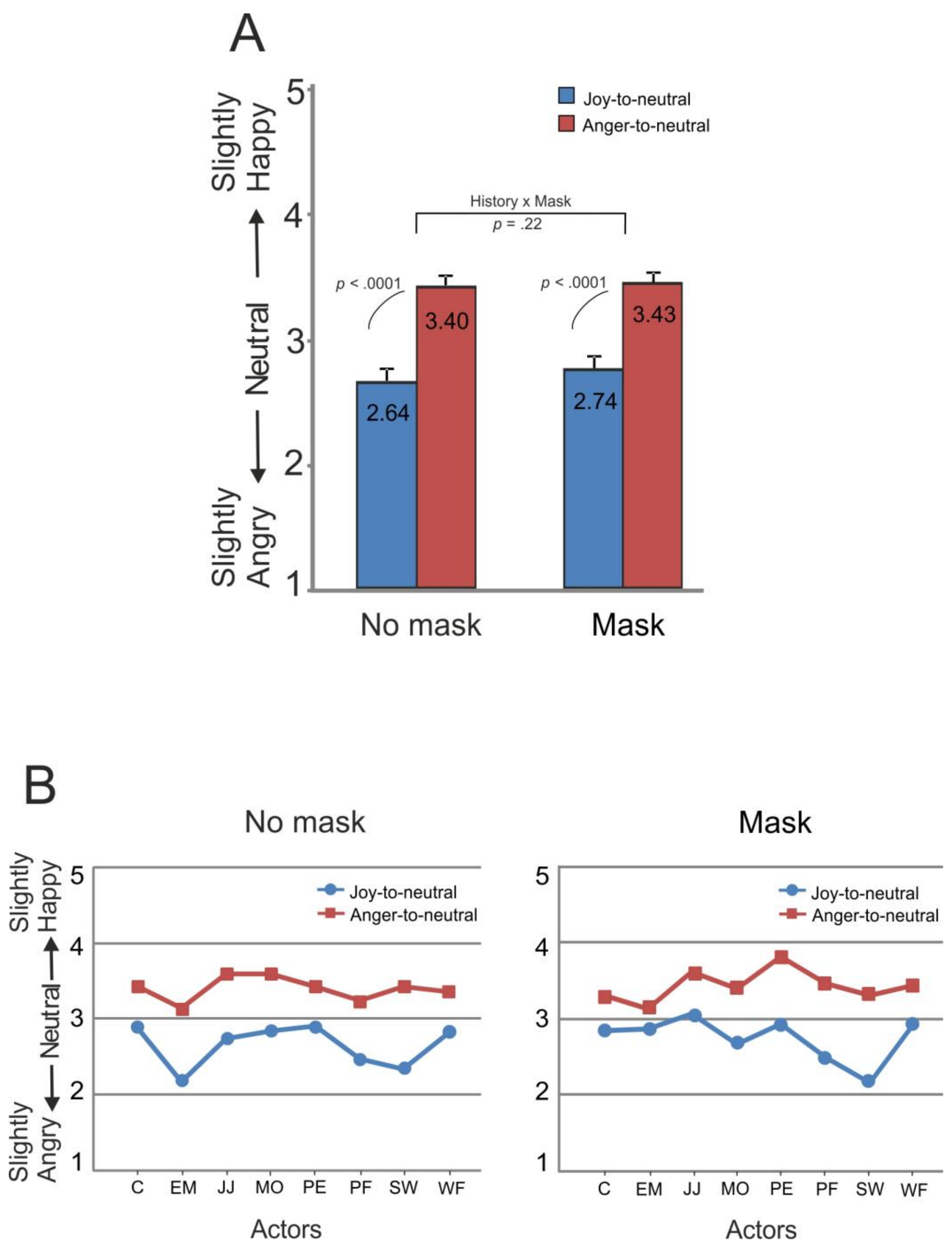

Figure 2.12. Results of Experiment 5. (A) Mean scores for the neutral expressions at the end of the Joy-to-neutral and Anger-to-neutral video-clips with an inserted mask (Mask) and without (No mask). Error bars indicate SEM. (B) Ratings for each of the eight actors in the two conditions (No mask and Mask) are shown to illustrate the consistency across actors.

2.6.3 Discussion

In Experiment 5 the flow of motion in the video-clips was interrupted by inserting a 400 ms long mask just before the last neutral frame (i.e. the frame that had to be evaluated) to specifically test for the role of extrapolation in bringing about the overshoot bias. The results showed that the overshoot bias survived the insertion of a 400 ms long mask; the extent of bias in the conditions with and without the mask was very similar. As this manipulation should have reset low-level mechanisms that rely on the flow of motion (cf. Enns et al., 2010), the conclusion is that such mechanisms did not play a crucial role in bringing about the overshoot bias.

It should be noted that the insertion of the mask induced quite an abrupt perceptual change. Yet, this sudden change in perceptual parameters did not affect the overshoot response bias in any way. This has also implications for the interpretation of the results of instant identity-change version of Experiment 2. That is, it shows that the rather abrupt change in low-level features associated with the transition between the two identities could not have caused the removal of the overshoot bias in that experiment. The mask in Experiment 5 introduced a considerably more abrupt change in low-level parameters than the instant transition between identities, yet the Mask experiment resulted in an overshoot bias whereas the instant identity-change experiment did not.

2.7 Experiment 6

In Experiment 6 the adaptation hypothesis was further investigated. Although the typical finding of adaptation studies is that adaptation periods shorter than 5 s do not produce any aftereffects, one study reported aftereffects at 500 ms duration of the adapting stimulus (Kovacs, Cziraki, Vidnyanszky, Schweinberger, & Greenlee, 2008). We therefore set out to test the adaptation hypothesis in more detail.

The possible contribution of adaptation to the overshoot response bias was tested by increasing the duration of the first frame of the video-clip to 600 ms, which meant that the full blown expressions (joy or anger) lasted the double amount of time as compared to the previous experiments. If the initial 300 ms duration of the first frame, despite its very short duration, was already able to induce an aftereffect, then doubling its duration may induce an increased after adaptation effect, which (if adaptation plays a role) should be reflected in an enhanced overshoot bias.

2.7.1 Method

2.7.1.1 Participants

Eighteen undergraduate Psychology students at Hull University took part. After applying exclusion criteria (see below), fourteen participants were included in the analysis (age, M = 19.4 years, SD = 1.2; 13 females, 1 male). None of the participants took part in any of the other experiments. All had normal or corrected-to-normal vision, gave informed consent, and received course credit for taking part.

2.7.1.2 Stimuli

The same Joy-to-neutral and Anger-to-neutral video-clips were presented as in the control condition of Experiment 5. In the experimental condition the duration of the first maximally happy or angry expression was increased to 600 ms. The total duration of the sequence was 1170 ms (Figure 2.13). In the control condition the duration of the first frame did not undergo any variation.

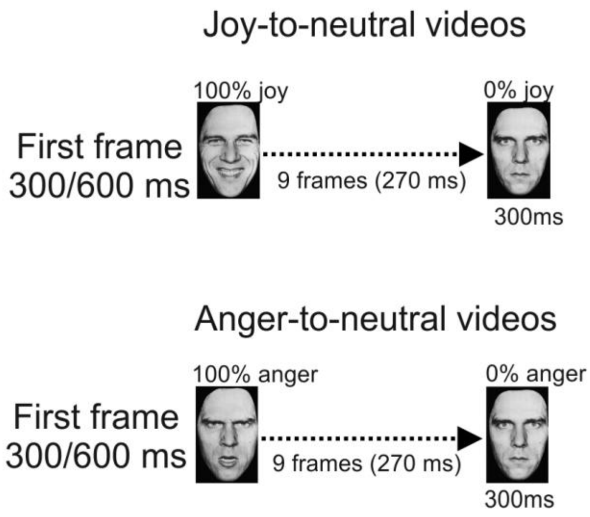

Figure 2.13. Schematic illustration of the stimulus presentations in Experiment 6. Joy-to-neutral videos (top panel) and Anger-to-neutral videos (bottom panel) were presented with two durations of the first frame (300 ms and 600 ms).

2.7.1.3 Experimental procedure

Participants first rated the static neutral expression of each actor, presented for 600 ms, by using the 5-point scale (ranging from slightly angry, via neutral, to slightly happy). Next, following 8 practice trials, two experimental blocks (one for the control and one for the experimental condition) were presented counterbalanced across participants. Each block consisted of 32 trials, (8 actors x 2 perceptual histories x 2 repetitions), presented in random order. The participants' task was to evaluate the last frame of the video-clips on the same 5-point scale as used in the calibration. Further procedures and apparatus were as in the previous experiments.

2.7.1.4 Data reduction and analysis

Exclusion criteria were as in the previous experiments. Trials in which the RTs fell below 250 ms or above 3000 ms were removed (3.8%). Participants were excluded if more than 25% of the RTs values were out of the mentioned range (n = 4). The ±2.5 SD rule to the mean difference ratings did not exclude any participant.

As in the previous experiments, all statistical analyses were performed on the calibrated scores.

2.7.2 Results

The results are shown in Figure 2.14. A 2X2 repeated measures ANOVA, with First frame (300 ms vs. 600 ms) and Perceptual history (Joy-to-neutral vs. Anger-to-neutral) as within-subject factors, showed a significant main effect for perceptual history (F(1, 13) = 85.97, p < .0001, η_p^2 = .869), but no significant main effect for the factor First frame (F(1, 13) = .087, p = .773, η_p^2 = .007. Importantly, no significant interaction effect was found (F(1, 13) = .006, p = .940, η_p^2 = .000). Thus, the degree of overshoot was independent of the duration of the first frame (300 ms or 600 ms).

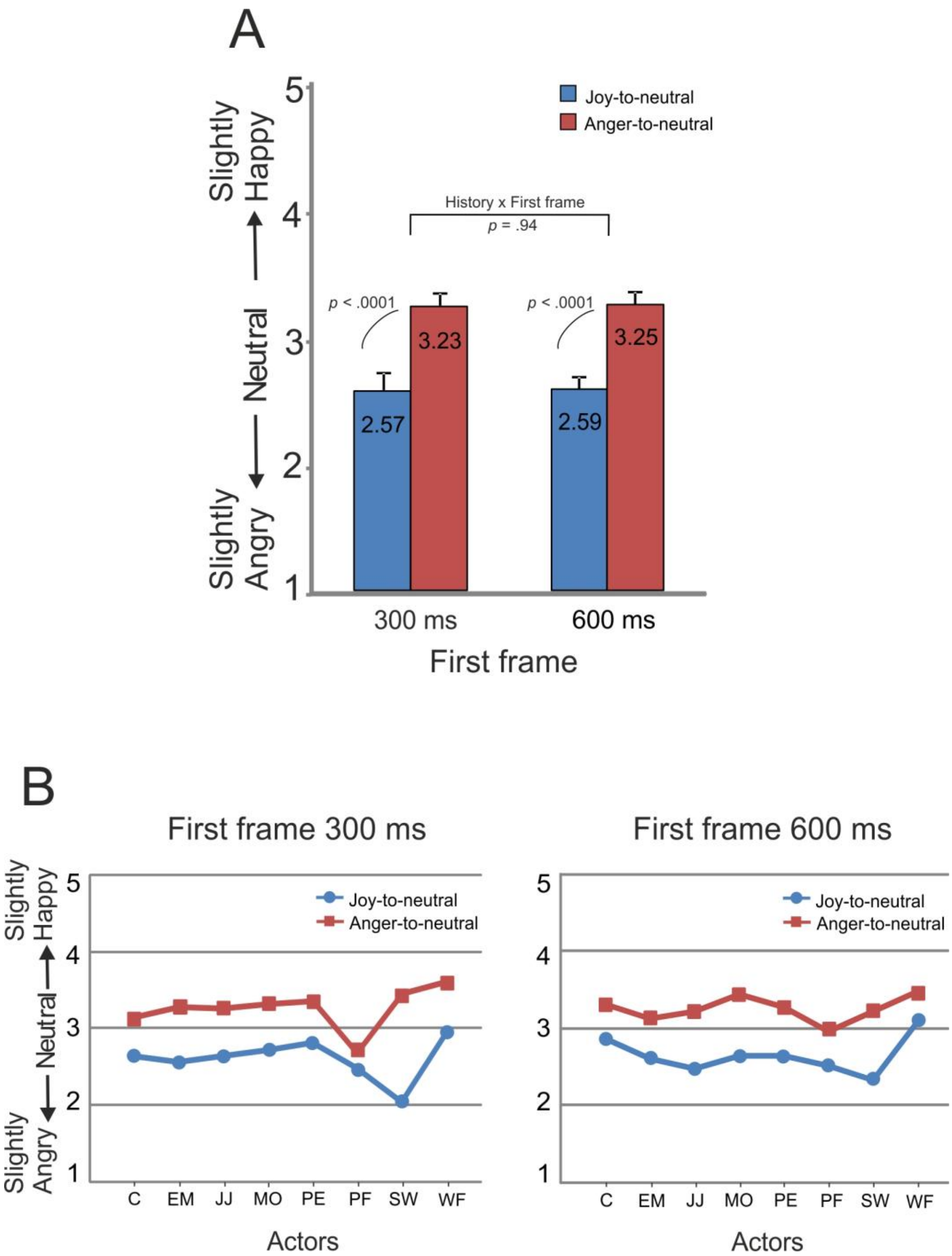

Figure 2.14. Results of Experiment 6. (A) Scores for the neutral expressions at the end of the Joy-to-neutral and Anger-to-neutral sequences. The first frame of the videos was presented for either 300 ms or 600 ms. Error bars indicate SEM. (B) Ratings for each of the eight actors in the two conditions (first frame: 300 ms and 600 ms) are shown to illustrate the consistency of the bias across actors.

2.7.3 Discussion

Experiment 6 was conducted to examine a possible contribution of visual adaptation to the overshoot response bias as found in the previous experiments. This was achieved by doubling the duration the first frame, depicting a maximally intense, expression. The rational was that if the first 300 ms frame was already able to induce an adaptation effect, then doubling its duration should have produced an enhanced overshoot bias. However, this manipulation did not modulate the amount of overshoot. A similar result was previously reported in Jellema et al., 2011, where the duration of the first frame was tripled as compared to the control condition. These findings, together with the finding of the identity change manipulation (Experiment 2), where the change of the actor's identity completely removed the response bias, strongly suggests that the overshoot response bias was not related to adaptation aftereffects. This is in line with reports that after-effects typically only occur after much longer exposures to the adapting stimulus and that they are not affected by identity variations (Leopold et al., 2001; Fox & Barton, 2007; Ellamil et al., 2008).

2.8 General discussion

The studies presented in Chapter 2 aimed at examining the contributions of basic bottom-up perceptual processes and top-down emotional anticipation to the generation of perceptual biases during the observation of dynamic emotional expressions, and the role these biases may play in anticipating and understanding other's emotional states.

In Experiment 1, the immediate, ecologically-valid and dynamic, perceptual history biased the perception of the neutral expression depicted at the end of the video-sequence. The identical neutral facial expression was consistently judged as more angry when presented at the end of Joy-to-neutral videos as compared to Anger-to-neutral videos. This bias was strong enough to overcome a 20% difference in emotional intensity. One interpretation of these results is that observers involuntarily kept track of the emotional state of mind of the actor and implicitly anticipated what the actor’s emotional state would most likely be in the immediate future. Such a top-down influence could bias the visual perception of the current facial expression in the direction of the anticipated emotional state of mind. We called this the ‘emotional anticipation’ hypothesis. However, a number of basic perceptual processes, such as sequential contrast/context effects, extrapolation or representational momentum (RM) and adaptation, might just as well be able to explain the bias in perceptual report. Therefore, in Experiments 2 to 6, the viability of the different explanations was tested.

In Experiment 2 a change in facial identity just before the final neutral expression was reached effectively removed the overshoot bias. This suggested that sequential contrast/context effects probably did not contribute to the bias, as the degree of contrast between the first and last frames of the videos remained unaffected by this manipulation. It also argued against an explanation in terms of an extrapolation on an underlying positive-to-negative valence dimension (RM) as the underlying valence dimension was also not affected by the manipulation.

Experiment 3 further assessed a contribution from sequential contrast effects. The last neutral frame was judged as slightly happy after the ‘joy loop’ and as slightly angry after the ‘anger loop’, which again argued against the sequential contrast and against the underlying positive-negative valence dimension, as both explanations would have predicted an overshoot bias.

The results of Experiment 4 argued against a contribution of RM effects, as such RM effects are typically modulated by the speed of the video-sequence (Yoshikawa & Sato, 2008), while we found that the speed variations did not affect the magnitude of the overshoot bias. In Experiment 5 the overshoot bias was found despite the insertion of a 400 ms long mask immediately before the last, neutral expression. This made clear that the overshoot bias did not depend on low-level visual mechanisms generated by the continuous flow of motion in the video-sequence, as such mechanisms would have been 'reset' by the mask insertion (cf. Enns et al., 2010). Finally, the possible role played by adaptation was specifically tested in Experiment 6. Here, doubling the duration of the first frame of the videos did not increase the overshoot effect, suggesting that the bias is not a typical adaptation aftereffect.

Taken together, these manipulations suggest that the overshoot bias found in Experiment 1 is not caused - at least not predominantly - by bottom-up perceptual mechanisms. A sequential contrast/context effect (cf. Suzuky & Cavanagh, 1998; Tanaka-Matsumi et al., 1995; Thayer, 1980) is an unlikely explanation as a change in identity removed the overshoot (Experiment 2), while a neutral starting-point did not result in a neutral evaluation of the endpoint, but in an undershoot (Experiment 3). RM operating on an underlying positive-negative valence dimension is an unlikely explanation for the overshoot bias because it would predict that a change in identity does not affect the bias yet it did (Experiment 2). Further, it would predict an overshoot in the 'Loop' experiment, whereas the opposite, i.e. an undershoot, was elicited (Experiment 3). The overshoot bias as found in the current study can also not be explained by RM operating on the dynamic facial features, because the speed of the videos did not modulate participants' responses (Experiment 4). Extrapolation of dynamic facial geometries can also be excluded as the insertion of a mask did not remove the overshoot (Experiment 5).

The explanation of RM operating on dynamic facial features can be further discredited by considering the extent of geometric change in the two perceptual histories. Happy facial expressions typically present a U-shaped mouth with raised cheeks and pulled up lip corners (Ekman & Friesen, 1975). Thus, in the happy-to-neutral morphs, the change of the U-shaped mouth (happy expression) into the flat mouth (neutral expression) might have been extrapolated in the observer's mind into a slightly inverted U-shaped mouth (denoting slight anger). However, the shape of the mouth in the maximally angry expressions was less well defined. Some actors (C, MO, PE, PF) showed a flat mouth as lips were pressed together while others (EM, JJ, SW, WF) displayed an open mouth with the lips parted in a square shape (Ekman & Friesen, 1975; Kohler et al., 2004). Neither a starting flat shape, or a starting square shape, would be extrapolated into a U shape typical of a smile at the end of the video-clip, yet in the Anger-to-neutral condition an equally large overshoot bias was found as in the Joy-to-neutral condition and the extent of the effect was consistent across actors. Changes in the height of the eye brows, which are typically lowered in angry expressions, can also not explain the response bias as in Joy-to-neutral morphs the eyebrow height hardly changed (cf. Kohler et al., 2004), yet a large overshoot bias was induced. Extrapolation of many other facial features and curves typically do not produce clear cut results.

Finally, adaptation can be ruled out as the change in identity should not affect the aftereffects induced by adaptation (Fox & Barton, 2007), at least not to a large extent, while this manipulation effectively removed the overshoot (Experiment 2). Adaptation was further excluded as a main contributor by Experiment 6, as doubling the duration of the first frame should have increased the overshoot bias if adaptation played a role in bringing about the bias, confirming earlier reports (Jellema et al., 2011).

As the above low-level visual mechanisms do not seem able to satisfactorily explain the overshoot response bias, it is suggested that instead an emotional anticipation mechanism was responsible for the perceptual biases. That is, the observer keeps track of any changes in the agent's emotional state of mind and continually updates its prediction, or anticipation, of what the most likely next emotional state of mind of the agent will be on the basis of the immediately preceding dynamics in the agent's facial expressions. These processes are proposed to be entirely involuntary.

2.8.1 Why would observers attribute to the agents a negative emotional/mental state in the Joy-to-neutral condition and a positive emotional/mental state in the Anger-to-neutral condition?

A recent study by Mühlberger, Wieser, Gerdes, Frey, Weyers & Pauli (2011) provided support for the contention that Joy-to-neutral clips induce a negative emotional/mental state attribution, while Anger-to-neutral clips produce a positive attribution. These authors presented clips depicting facial expressions that morphed between neutral and either joy or anger, in both directions. Using fMRI, they found that the joy-to-neutral and neutral-to-joy clips activated largely different brain areas, and similarly that the anger-to-neutral and neutral-to-anger clips also activated largely different brain areas. The joy-to-neutral clips shared a strong common activation with the neutral-to-anger clips in the lateral orbitofrontal cortex bilaterally, and in the left amygdala and left insula. The anger-to-neutral clips shared a strong common activation with the neutral-to-joy clips in the left dorsal striatum. This suggests that the joy-to-neutral clips are to some extent processed similarly as neutral-to-anger clips, but quite differently from neutral-to-joy clips. Their common denominator may be their negative/threatening character, resulting in activation of threat-related areas such as the amygdala (Morris et al., 1996; Whalen et al., 1998) and the insula (Schienle et al., 2002). It also suggests that the anger-to-neutral clips are to some extent processed similarly as neutral-to-joy clips, but quite differently from neutral-to-anger clips. For them, the common denominator may be their positive/rewarding character, reflected by activation of reward-related areas, such as the ventral striatum and putamen (Morris et al., 1996; Whalen et al., 1998). Watching an angry person calming down may serve as a positive reinforcer, activating reward circuitries (Mühlberger et al., 2011). Thus, the onset of a happy expression and the offset of an angry expression share a positive valence, while the offset of a happy expression and the onset of an angry expression share a negative valence.

Thus, one possibility is that the anger-to-neutral clips activated motivational areas in the brain related to approach and/or reward, even though the anger expression itself would activate avoidance and/or threat areas (Öhman, 1993; Whalen et al., 1998). The joy-to-neutral clips may have activated areas related to avoidance and/or threat, even though the joy expression on its own would activate approach and/or reward areas (Whalen et al., 1998). Alternatively, or in addition to such an activation of motivational areas, the joy/anger-to-neutral clips could activate areas involved in simulation such as the premotor cortex and parts of the parietal cortex (Gallese et al., 1996), where mirror mechanisms might enable the observer to understand the dynamic emotional action and to anticipate the most likely next emotional state, on the basis of the observer's own experiences (Gallese, Keysers, & Rizzolatti, 2004; Goldman, 2006).

2.8.2 Possible mechanisms underpinning emotional anticipation

Widely different models have been proposed for the way in which we understand and anticipate other's emotional/mental states. The two main approaches to intersubjective understanding are often referred to as Theory Theory (TT) and Simulation Theory (ST). According to TT the attribution of emotional/mental states reflects the theoretical use of folk psychology, tacit knowledge and psychological laws (e.g. Fodor, 1968; Leslie, 1994b). ST might involve either an explicit simulation (i.e. putting oneself in the other's shoes; e.g. Gordon, 2004), or an implicit (subpersonal) or 'embodied simulation', (i.e. simulating others' actions and emotions in one's own sensorimotor system (e.g. Gallese, 2007). ST received a huge boost by the discovery of mirror neuron systems (Gallese et al., 1996; Rizzolatti & Craighero, 2004), which are proposed to play a crucial role in the simulation process. In addition, some authors proposed hybrid approaches, in which the immediate and involuntary understanding of, in particular, other's emotions is subserved by mirror mechanisms and embodied simulation, while more sophisticated intersubjective understanding is guided by TT principles (e.g. Goldman, 2006).

In parallel, researchers in neuroscience focused their attention on the neural mechanisms underlying the perception of faces and the emotional expressions they convey, and the ways in which these stimuli affect the understanding of the agent's emotional state of mind. Single unit studies in the macaque monkey (see Jellema &

Perrett, 2002a, b; 2005, for reviews) and fMRI studies in humans (e.g. Haxby, Hoffman, & Gobbini, 2002) have shown that especially the processing of the variant aspects of faces, including their emotional expressions, involves the superior temporal sulcus (STS). STS cell responses to the image of a face were shown to depend on the immediate perceptual history of that face image and how it is embedded in a natural action sequence (Jellema & Perrett, 2003; Perrett et al., 2009). The STS therefore may well be implicated in the bias in perceptual report found in the current study. Although STS cells might generate expectations, or anticipations of impending behaviour of others (Jellema & Perrett, 2003; Perrett et al., 2009) and may play a role in processing other's intentions (Jellema & Perrett, 2000; Saxe, 2006; Saxe et al., 2004), intuitive and immediate understanding is unlikely to be accomplished by just the STS. The description of currently observed actions generated by the STS may activate its motor equivalent in the MN areas, which in turn generates associated internal states, i.e. those states that would have been evoked would the observer herself have executed that particular action. These states are then attributed to the actor. Hence, a distributed system, involving the STS and the mirror neuron system, might underlie the tacit/intuitive understanding of others' facial expressions through motor simulation.

This might (top-down) shape the perception of the visual stimulus as reported here, without requiring any conceptual reasoning (cf. Gallese & Goldman, 1998). The hypothesised neurophysiological correlates of emotional anticipation are in harmony with various accounts of the involvement of embodiment in the perception of facial expressions (Carr et al., 2003; Niedenthal, 2007; Pitcher et al., 2008, Wicker et al., 2003).

2.8.3 Limitations of the current studies and future research directions

The studies in Chapter 2 present some limitations. First, only happy and angry facial expressions were used. This choice was motivated by the requirement to have opposite expressions in terms of positive-negative valence but fairly similar expressions in terms of arousal (Russell & Bullock, 1985). Thus, at present the emotional anticipation hypothesis cannot be extended to other emotional facial displays and further examinations are needed. Second, even though the results found in the six experiments seem to rule out a bottom-up explanation based on contrast effects, adaptation, and RM for the overshoot response bias, they do not provide *direct* evidence for a top down affective account. It would be interesting to further test the emotional anticipation hypothesis using deceptive facial expressions (Ekman, 2003), which, despite their similarities with genuine expressions in terms of features and geometries, do not have the same link with the emotional state of the agent.

Another possible line of research could involve the use of electromyographical measures (e.g. Dimberg, Thunberg, & Elmehed, 2000) of the observer's facial muscles to investigate the role of motor resonance and embodied simulation. Specifically, it would be relevant to test whether at the end of the Joy-to-neutral video clips EMG activity occurs in the observer's 'frowning' muscle (Corrugator Supercilii) and no activity in the observer's 'smile' muscle (Zygomaticus Major). This would be surprising as in these video-clips there is no angry facial expression at all. Similarly, one could test whether at the end of the Anger-to-neutral video clips EMG activity occurs in the observer's Zygomaticus and no activity in the observer's Corrugator. Again, that would be surprising as in these video-clips there is no happy facial expression at all. If these hypotheses are confirmed then that would provide compelling evidence for embodied simulation as the underlying mechanism for the perceptual overshoot bias.

Further investigations should also try to disentangle the specific contributions of embodied simulation and of a 'tacit' version of TT in emotional anticipation. The automatic and involuntary nature of the processes does not need to constitute an obstacle for TT as folk psychological theory is partly 'tacit' (Fodor, 1968; Stich & Nichols, 1992).

2.8.4 Conclusions

The perception of neutral facial expressions in this first experiment was not biased by basic visual processes; rather it seems that it was shaped by the attribution of the anticipated emotional state of mind of the agent, based on the immediate dynamic perceptual history of the face. These findings are in line with the interactive model of social perception, recently proposed by Teufel and colleagues (Teufel et al., 2010), which emphasises the influence of mental attributions to conscious social perception.

According to theories of embodied cognition, the recognition of facial expressions is not only the result of visual processing, but also involves motor simulation of the observed expression and/or activation of associated somatovisceral responses (Gallese et al., 1996; Rizzolatti et al., 1996; Adolphs 2000; Goldman, 2006; Gallese, 2007). Several imaging studies provided support for the role played by embodied mechanisms in emotional face processing, mediated by mirror mechanisms (Carr et al., 2003; Niedenthal, 2007; Pitcher et al., 2008; Wicker et al., 2003). In addition, simulation of observed actions has been argued to not only influence the recognition of the action, but may also facilitate the understanding of the action in terms of intentions and goals (Gallese, 2007; Ortigue, Sinigaglia, Rizzolatti, & Grafton, 2010). In line with this way of thinking, the recently proposed ‘simulation of smiles model’ (SIMS) by Niedenthal, Mermillod, Maringer & Hess (2010) suggests that the observer’s motor experience in response to the perception of a facial expression facilitates the understanding of the emotional state of the agent. These involuntary simulation processes might enable a form of ‘low-level’ emotion reading, which does not require any deliberate internal reasoning (Goldman, 2006; Goldman & Sripada 2005).

The emotional anticipation hypothesis may reflect the activation of mirror mechanisms triggered by the visual information in the perceptual history, which implicitly enables the anticipation of the other’s future emotional state of mind. In this view, emotional anticipation does not rely on the use of a ‘theory’ as in Theory of Mind, but occurs implicitly and automatically (cf. Goldman, 2006).

The overshoot bias found in our studies does not seem to arise from high-level theorising, a deliberative use of psychological rules, or explicit knowledge. Rather, it is proposed that the perceptual history gives rise, in an involuntary manner, to ‘emotional anticipation’, i.e. a type of ‘low-level mindreading’ (cf. Goldman, 2006; Goldman & Sripada, 2005), which in turn affects perception. However, direct evidence for this interpretation is not yet available at this stage (but see ‘Future research directions’ for specific additional investigations to shed light on the possible contribution of a simulation’s account).

At present, the emotional anticipation hypothesis with its underlying embodied simulation mechanism as explanation for the response bias found in the current study are still highly speculative. It would be relevant to test the emotional anticipation task in clinical populations with impairments in emotion reading and social/communicative domain, such as individuals with Autism spectrum disorders (ASD), as this might shed light on the possible underlying mechanisms. This is what will be done in the next chapter.

CHAPTER 3

Emotional anticipation in adults with Asperger's syndrome

3.1 Introduction

Chapter 2 reported that a rapid decrease of the intensity of emotional facial expressions, i.e. a happy or angry expression that gradually morphed into a neutral expression, biased the perception of that last neutral expression in individuals with typical development (TD; Jellema et al., 2011). This perceptual distortion involved a crossing of the category boundary: the neutral expression at the end of the Joy-to-neutral video-clips was evaluated as slightly angry, and the same neutral expression at the end of the Anger-to-neutral video-clips was evaluated as slightly happy (‘overshoot’ response bias). In Chapter 2, a number of investigations, aiming at assessing the underlying mechanisms responsible for this phenomenon, were reported. Specifically, these investigations tried to disentangle the roles played by bottom-up visual perceptual processes (such as contrast effects, adaptation and extrapolation or Representational Momentum), and a top-down mechanism called ‘emotional anticipation’. The findings suggested that the perceptual history immediately preceding the last neutral expression gave rise to an involuntary anticipation of the emotional state of the agent, which in turn gave rise to the perceptual distortion of that neutral expression, reflected by the overshoot response bias. Emotional anticipation was described as a low-level (automatic and involuntary) mind-reading mechanism, which enables the tacit/intuitive understanding of others’ emotional states of mind. Debriefing suggested that TD individuals performed the tasks without adopting any conscious/cognitive strategy involving deliberative reasoning.

The importance of an emotional anticipation mechanism is clear from day-to-day interactions. Our social life would indeed be extremely difficult if every time we interact with others we would need to effortfully infer the reasons underlying their behaviour. Rather, we immediately grasp others' emotional and mental states from facial and/or bodily expressions and we keep track of someone's intentions by monitoring eye gaze direction or meaningful gestures. The ability to 'read' social cues automatically and involuntary in terms of the agent's inner states is essential to make sense of - and be part of - the social world. TD Individuals are equipped with such intersubjective understanding or Theory of Mind (ToM; Baron-Cohen, 1995). Importantly, it enables one to *anticipate* the behavior and mental/emotional states of others, thus promoting appropriate social interactions.

Whenever we perceive another person performing a certain act (including adopting a static posture), we tend to attribute a mental and/or emotional state to that person. This compelling tendency of typically-developed humans to attribute mental/emotional states to others, or to 'mindread' others, has been well documented (e.g. Baron-Cohen, 1995). The premier source of information on the basis of which such attributions are formed is directly derived from the (visual) perception of the other's body. It consists of bodily cues such as gaze direction, facial expressions, postures and action sequences, often in relation to objects in the environment (context).

The bottom-up pathways along which the 'perceptual' processing of these social stimuli/cues take place are fairly well understood (Allison et al., 2003; Puce & Perrett, 2003). It should be noted that by 'perceptual' processing of a social stimulus we mean the processing in terms of physical or geometrical features and dynamics of the stimulus, possibly in relation to other cues, but devoid of any 'higher' psychological meaning. This constitutes what one could call a 'mechanistic' description (Jellema & Perrett, 2002a, b; 2005). It is generally accepted that these mechanistic descriptions of social cues allow one to form inferences about the emotional/mental state of the agent (Jellema & Perrett, 2005; Blakemore & Decety, 2001). These attributions tend to be susceptible to prior explicit knowledge about the other, acquired by the observer at an earlier stage. A plethora of studies show that when confronted with one and the same action, our attributions will vary depending on the nature of the prior knowledge (or prejudices) (Gamond et al., 2011; Kozak, Marsh, & Wegner, 2006; Read, Druian, & Miller, 1989). This description of social cue processing reflects a fairly rigid, fixed processing protocol. The top-down modulation of the visual perception kicks in relatively late, i.e. after the social cues have been analysed and attributions on the basis of these cues have been formed, possibly modulated by prior knowledge and context.

However, recently studies have challenged the dogma that there is a marked distinction between ‘perceptual’ processing on the one hand and ‘mentalising’ on the other hand. For example, Teufel, Fletcher, & Davis, (2010) build on the idea of a bidirectional relationship between the mechanisms supporting the perceptual processing stages, and the attributions made by the observer about the agent’s inner states on the basis of prior explicit knowledge. Teufel et al. (2010) made the point that low-level ‘perceptual’ processing of social cues is already susceptible to explicit knowledge about the actor’s mental state and can be modulated according to such attributions. They referred to this interaction with the term ‘perceptual mentalising’.

We extend this notion of an interactive relationship by proposing that the perceptual processing of social cues also interacts with attributions made on the basis of the immediate perceptual history of the observed action (Palumbo & Jellema, under review). These latter attributions are thought to reflect the operations of a prediction or anticipation mechanism, which operates automatically and involuntarily, and which, in a way, is part of the perceptual system. In Teufel et al.’s model the observer is fully aware of the attributions, as they reflect explicit knowledge provided by the experimenter to the observer prior to the task. In our model, the processing of the dynamic facial expressions generates in an automatic/involuntary fashion a prediction or anticipation in the observer about what the actor's most likely next mental/emotional state of mind will be. This happens 'on line' during the task, whereby the anticipation of the most likely next state of mind is continuously updated on the basis of the immediately preceding events, without the observer being aware of it. These ideas blur the distinction between perception and anticipation (or mentalising) as it basically incorporates anticipation (or mentalising) within the perceptual process. That is, anticipation is an automatic consequence of the dynamic perceptual input stream. Thus, the mental/emotional anticipations do not involve deliberate reasoning, but can be entirely involuntary and automatic. It is as if the mere perception of the social stimulus automatically and subconsciously induces 'mentalizing' activities, which then in turn modulate the low-level bottom-up perception. It enables the observer to implicitly, or tacitly, form expectations and anticipations about the most likely next action and/or state of mind (cf. Hudson et al., 2009; Hudson & Jellema, 2011). Similarly, the literal or physical emotional expressivity of a face, especially when it is embedded in a dynamic sequence, may automatically lead to an anticipation of what the agent's emotional state most likely will be in the immediate future, which then affects the perceptual judgments (Palumbo & Jellema, under review).

3.1.1 Social cue processing and 'emotion-reading' in ASD

Autism Spectrum disorder (ASD) is a pervasive developmental condition characterized by impaired social development, dysfunctional communicative skills, and stereotypical, repetitive behaviours, which are often accompanied by obsessive interests and a lack of emotion, empathy and pretend play (Rutter, 1978; WHO, 1992; DSM-IV, 2004; Baron-Cohen, 2005). Despite these marked deficiencies, many other cognitive abilities are intact (Happé and Frith, 2006).

Under the ASD umbrella, Asperger's syndrome (AS) is a relatively mild form of autism, in which the development of language and cognitive skills is preserved, while IQ scores are in the normal, or above normal, range. The distinction with high-functioning autism (HFA), in which there is a delay in the development of language and cognitive abilities, is often difficult to make as records of symptoms in early development may be absent. In addition to mild problems in the communicative and relational domains, individuals with AS may also show limited empathy, lack of social or emotional reciprocity, and impaired nonverbal behaviors related to eye contact, facial expression, posture, and gesture (McPartland & Klin, 2006). These difficulties have been linked to a failure to attribute epistemic mental states (such as desires, beliefs and intentions) to others (Frith et al. 1991; Baron-Cohen, 1995; Baron-Cohen et al., 1995) and to a failure to involuntary or spontaneously interpret others' social cues (Senju et al., 2009; Jellema et la., 2009).

Even though individuals with HFA or AS may fail to compute others' mental states spontaneously, they can use explicit (non-social) information to try and understand others (Kuchinke et al., 2011; Kuzmanovic et al., 2011; Senju et al., 2009). Thus, through explicit reasoning about others' inner states, individuals with AS may compensate for an automatic failure, although this would render their social interactions more slow and effortful.

3.1.2 Anticipation in individuals with AS

Recently, it was suggested that individuals with AS specifically have problems with *anticipating* others' actions (Zalla et al., 2010; Hudson et al., 2011). These deficits may, however, be compensated by alternative strategies, such as relying on physical dynamic characteristics of the agent (Hudson et al., 2011). In this study by Hudson et al. (2011), TD and AS participants believed that the head of an agent had rotated further when the agent's gaze was leading the direction of rotation than when it was lagging behind rotation. However, while estimations in the TD group were influenced by the involuntary attribution of intentions to the agent conveyed by the eye gaze direction, in the AS group the action anticipation seemed to be caused by the low-level visual appearance of the gaze direction rather than by the inferred intentions. This latter contention was suggested by the finding that the AS group was influenced to the same extent by the gaze of the agent as by the 'equivalent of gaze' in an inanimate cylindrical object. It remains an open question whether the inability to involuntary anticipate others' actions in AS individuals is caused by an inability to keep track of the agent's state of mind. It is also unknown to what extent they may use atypical means to compensate their failure.

3.1.3 Emotional face processing in ASD

The extent to which individuals with ASD are able to recognize and understand emotional facial expressions is still much debated despite a large amount of research into this topic. This is mainly due to a large variety of paradigms and stimuli (usually static pictures which are less ecologically valid), different age and different levels of severity of ASD. On the one hand there is evidence about ASD deficits in processing emotional facial expressions (e.g. Celani et al., 1999; Corden et al., 1994; Grossman, Klin, Carter, & Volkmar, 2000; Hobson, 1986a, 1986b; Humphreys, Minshew, Leonard, & Behrmann, 2007; Law Smith et al., 2010; Lindner & Rosen, 2006; O' Connor, Hamm, & Kirk, 2005; Tantam et al., 1989; Teunisse & de Gelder, 2001), while on the other hand some studies show that individuals with ASD are relatively unimpaired in processing emotional facial expressions (e.g. Adolphs et al., 2001; Baron-Cohen et al., 1997; Castelli, 2005; Gepner et al., 2001; Loveland et al., 1997; Ogai et al., 2003; Ozonoff et al., 1990; Prior et al., 1990; Wicker et al., 2008; Wong et al., 2008).

According to the Weak Central Coherence theory (WCC; Happé, 1999; Happé and Frith, 2006), social perception in individuals with ASD is impaired because they focus their attention on the local details of a stimulus, rather than on the whole. Therefore, ASD individuals may analyze facial expressions on the basis of isolated elements, or low-level features (e.g. the U shaped mouth in a smile), thereby losing sight of the global configuration and the meaning of it. Another possibility is that individuals with ASD, due to their difficulty in ascribing intentions to animate objects via referential cues, may adopt systemizing mechanisms (i.e. based on input-operation-output relations) to understand and anticipate other people's behaviour (Baron Cohen, 2002; Hudson et al., 2011; Ristic et al., 2005). This might also account for their lack of empathy (Baron-Cohen, 2006). The prevalence of a mechanistic processing mode, as opposed to a mentalistic one (Driver et al., 1999; Jellema & Perrett, 2002a, 2002b; 2007), may also contribute to the impairments in social perception in ASD.

3.1.4 The current study

In Chapter 2 the study of TD individuals revealed that the perceptual judgment of a face with neutral expression is biased (overshoot effect) by the immediate perceptual history. It was suggested that the underlying mechanism consisted of an involuntary anticipation of the agent's emotional state of mind, generated by the perceptual history immediately preceding that neutral expression (Palumbo & Jellema, under review).

The experiments in Chapter 3 aimed at testing the same 'emotional anticipation' task on individuals with AS, matched with TD control groups. In Experiment 7 the question was whether individuals with AS would also show an overshoot response bias. The hypothesis was that they would not engage in 'emotional anticipation', in accordance with theories about their assumed lack of empathy and ToM (Baron-Cohen, 1995). However, on the basis of previous findings suggesting that AS individuals can anticipate someone else's actions by using alternative strategies (Hudson et al., 2011), it might be that AS individuals would produce an overshoot bias in their judgments, but through a different route.

3.1.4.1 The recruitment of the clinical group

The clinical group consisted of twenty-five students recruited through disability services at universities in the North-East of England (UK). All these students had previously received a diagnosis of AS, based on DSM-IV criteria (American Psychiatric Association, 1994). The diagnostic evaluations involved a review of prior records/psychiatric observation. For the purpose of the current experiments the ADOS (Autism Diagnostic Observation Schedule, module 4; Lord, Rutter, DiLavore, & Risi, 1999), was administered by a qualified psychologist. The ADOS is a semistructured, standardized interview using developmentally appropriate social interactions and equipment. Algorithms are calculated for social interaction, communication and repetitive behaviours. An ADOS score of 7 or higher is indicative of clinical levels of ASD. All participants in the clinical group also completed the Autism Spectrum Quotient questionnaire (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001; Hoekstra, Bartels, Cath, & Boomsma, 2008; Wheelwright et al., 2006). The AQ is a fifty-statement, self-administered questionnaire, designed to measure the degree to which an adult with normal intelligence possesses autistic-like traits. It covers social skills, attention switching, and attention to detail, communication and imagination.

In the clinical group three participants with ADOS score below 7 were excluded as they did not meet the classification of AS (these participants were not tested in any of the experiments presented in this chapter). Therefore, the final clinical group was constituted of twenty-two AS individual (see participants' individual characteristics in Table 3.1).

Table 3.1. The clinical group: individual characteristics. Age (years); Gender (F, M); AQ, Autism-spectrum Quotient; IQ-T, total IQ score; IQ-V, verbal IQ score; IQ-P, Performance IQ score; ADOS, autism diagnostic observation schedule. All participants' native language was English. Participants highlighted in grey were excluded because they had ADOS scores < 7 (they were not included in the calculations of the means).

Participants	Age	Gender	AQ	IQ-TOT	IQ-V	IQ-P	ADOS
1	19	M	18	122	113	134	7
2	19	M	20	121	130	107	9
3	18	F	36	110	107	113	9
4	21	M	22	95	97	91	9
5	22	M	40	140	142	127	7
6	21	F	35	122	107	142	9
7	44	F	40	120	125	109	9
8	18	M	20	101	96	107	9
9	19	M	44	107	110	104	7
10	31	F	25	119	116	119	7
11	20	M	28	124	114	134	9
12	28	M	34	117	120	111	13
13	19	M	30	117	105	125	7
14	19	F	31	117	125	105	7
15	21	M	33	117	125	105	8
16	21	M	39	128	128	121	9
17	18	M	37	116	144	84	7
18	18	M	12	117	122	109	8
19	33	M	34	113	115	102	8
20	19	M	35	110	114	105	8
21	19	F	38	125	138	106	7
22	18	M	30	122	125	113	7
23	21	F	28	123	122	121	5
24	19	M	21	104	107	100	4
25	20	F	29	122	125	113	6
MEAN	22.0	6 famales	31	117	119	112	8
SD	6.4	16 males	8.3	9.3	13.0	13.9	1.4
N	22		22	22	22	22	22
SE	1.4		1.8	2.0	2.8	3.0	0.3

All 22 individuals performed the Control condition. However, to avoid familiarization effects, only in two experiments (Experiments 7 and 9) the Control condition was included, while in the other two experiments (8 and 10) it was absent. For AS participants in Experiments 8 and 10, their scores obtained in the Control condition of either Experiment 7 or 9 were used in the analyses.

3.2 Experiment 7

Short video-clips of dynamic facial expressions, which morphed from 100% joy or anger to a 0% neutral expression (and to a 10% joy or a 10% anger expression), were presented to adults with AS, and to a control group consisting of TD participants matched for age, gender and intellectual abilities. Participants were asked to judge the last expression of the sequence, which last expression was presented for 300 ms. The primary hypothesis was that the individuals with AS do not engage in ‘emotional anticipation’ and therefore will not show an overshoot bias in their evaluations.

However, in principle, there was also the possibility that they would compensate for a lack of ‘emotional anticipation’ by using low-level visual effects, such as contrast effects or extrapolation to solve the task. In the control group, it was predicted to replicate the findings of Experiment 1 (Chapter 2).

3.2.1 Method

3.2.1.1 Participants

All AS and TD participants had normal or corrected-to-normal vision, and provided written consent prior to the experiment. Participants received course credits or a fee for taking part. The study was approved by the Ethics committee of the Department of Psychology of Hull University.

AS group

AS participants who took part in Expriment 7 consisted of seventeen students. After applying exclusion criteria (see Data reduction below for details) one student was excluded from the analysis. The sixteen students included (5 females, 11 males; mean age = 20.3 years, SD = 3.2) had a mean total ADOS score of 8.0 (SD = 0.9) and a mean AQ score of 32.1 (SD = 7.4). Their mean total IQ score was 117.1 (SD = 10.9; assessed using the Wechsler Adult Intelligence Scale (WAIS-III, Wechsler, 1997).

TD group

The control group consisted of twenty undergraduate Psychology students at Hull University (UK). The TD group was recruited to match the AS group for age, gender and intellectual abilities. After applying exclusion criteria (see Data reduction below for details), the data of 18 participants were included in the analysis (age, M = 20.1 years, SD = 4.0 years; 7 females, 11 males). All individuals in this control group completed the AQ. Their mean AQ score was 15.3 (*SD* = 5.2). Their mean total IQ score was 113.4 (*SD* = 8.3); see Table 3.2.

The AS and TD groups did not differ significantly in terms of age ($t(32) = .554$, $p = .583$), gender ratio ($X^2(1, 33) = .216$, $p = .642$) nor IQ ($t(32) = -1.14$, $p = .264$). The AQ scores were significantly higher in the AS group than in the TD group ($t(32) = -7.74$, $p < .0001$)

Table 3.2. AS and TD groups in Experiment 7: participant characteristics. Age is in years. Standard deviations are shown between brackets. F, female; M, male; AQ, Autism-spectrum Quotient; IQ-T, total IQ score; IQ-V, verbal IQ score; IQ-P, Performance IQ score; ADOS, autism diagnostic observation schedule.

Group	n	Age	Sex	AQ	IQ-TOT	IQ-V	IQ-P	ADOS
AS	16	20.3 (3.2)	5 F; 11 M	32.1 (7.4)	117.1 (10.9)	119.9 (14.7)	111.4 (14.7)	8.0 (0.9)
TD	18	20.1 (4.0)	7 F; 11 M	15.3 (5.2)	113.4 (8.3)	114.1 (7.6)	111.8 (11.6)	

3.2.1.2 Stimuli and experimental procedure

The stimuli and experimental procedure were the same as in Experiment 1 of Chapter 2. In short, the stimuli consisted of video-clips depicting a maximally happy or angry expression of which the intensity gradually decreased until a neutral (or a 10% joy or 10% anger) expression was reached. The first and the last frames lasted 300 ms each.

The nine interpolated frames between the full-blow expression and the last frame lasted 270 ms (9 x 30 ms). The interpolated sequence was 30 ms longer or shorter for clips with 10% intensity endpoints (see Figure 3.1). An illustration of the stimuli is provided in Figure 3.1.

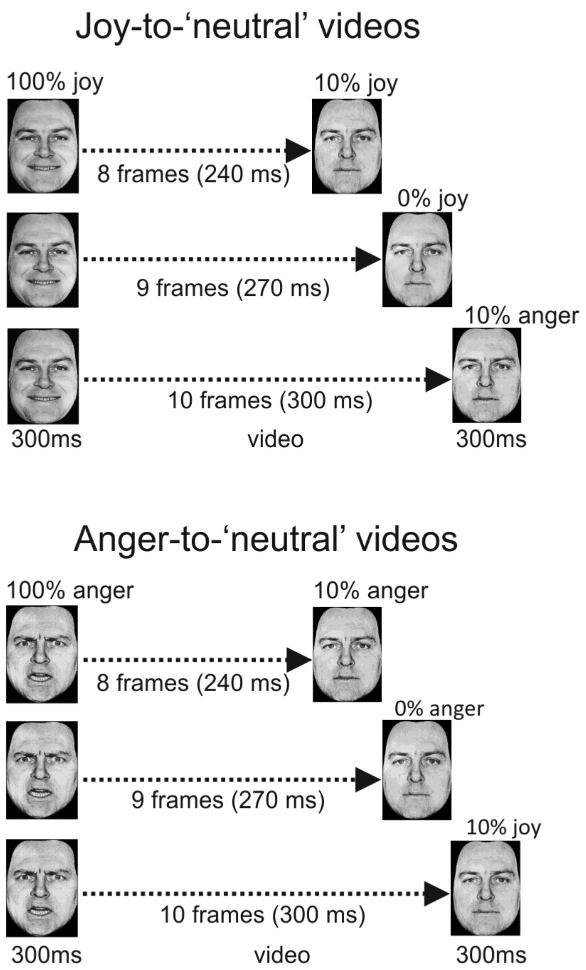

Figure 3.1. Schematic illustration of the stimulus presentations in Experiment 7. Joy-to-'neutral' videos started with a facial expression of intense joy, which gradually morphed into a 10% joy, a neutral or a 10% anger expression (top panel). Anger-to-'neutral' videos started with a facial expression of intense anger, which gradually morphed into a 10% anger, a neutral or a 10% joy expression (bottom panel).

3.2.1.3 Experimental design

Participants first rated the neutral expression of each actor presented in isolation as in Experiment 1 (16 trials) for the purpose of calibration (see below). Then, following 8 practice trials, 64 randomized experimental trials were presented, half of which started with a happy expression (called Joy-to-‘neutral’ videos) and half with an angry expression (called Anger-to-‘neutral’ videos). Of both types of videos, 50% of trials ended with the neutral expression, 25% of trials ended with the 10% joy and 25% of trials with the 10% anger expression.

3.2.1.4 Data reduction and analysis

Trials in which RTs were below 250 ms or above 3000 ms were considered outliers and were removed (3.8% of total number of trials). Participants were excluded if more than 25% of the RTs values were outside the above range (n = 0). A $\pm$2.5 SD rule was applied to the mean difference ratings (i.e. rating in the Anger-to-neutral condition minus rating in the Joy-to-neutral condition), which excluded 1 participant. Further, participants were removed if their judgments on the scale were made by pressing the same key for more than 90% of the total number of trials (n = 0).

In the control group the same criteria were applied. 2.9% of total number of trials were removed because RTs were below 250 ms or above 3000 ms. One participant was excluded for having more than 25% of RTs values outside the above range. The $\pm$2.5 SD rule to the mean difference ratings excluded one more participant.

In both groups the calibration scores were used to adjust the scores in the experimental trials; a calibration factor [calibration factor = 3.00 minus the calibration score] was added to the experimental scores. An illustration of the calibration scores for the neutral expressions of each actor in AS and TD individuals is provided in Figure 3.2. Note that the calibration scores were fairly similar across the two groups (and also across the experiments in this chapter), with the neutral expression of actors C and WF systematically rated as slightly angry (see Figure 3.2). The calibration procedure allowed to perform one-sample t tests with test value 3.00 on the scores in the two perceptual histories. All statistical analyses were performed on the calibrated scores.

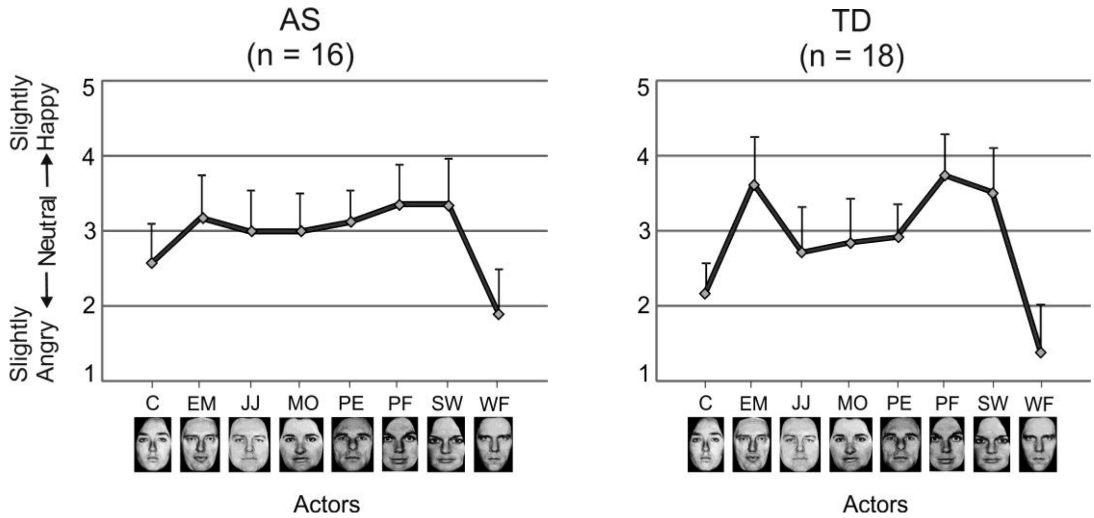

Figure 3.2. Calibration scores in Experiment 7. Ratings on the 5-point scale (vertical axis) for the neutral expression presented in isolation. Scores are shown for each of the 8 actors (their initials are indicated on the horizontal axis) for AS (left panel) and TD individuals (right panel). Error bars indicate +1SD.

3.2.2 Results

The results are illustrated in Figure 3.3 (for clarity the TD and AS groups are shown in separate graphs). A 3x2X2 mixed ANOVA was performed with as within-subject factors Endpoint of the video (10% anger vs. neutral vs. 10% joy) and Perceptual history (Joy-to-'neutral' videos vs. Anger-to-'neutral' videos), and as the between-subjects factor Group (AS vs. TD). The main effect of the Endpoint factor was significant (F(2, 64) = 43.9, p < .0001, η_p^{2} = .58), reflecting that when the videos ended at 10% joy the expressions in the last frame were judged as more happy then when they ended at 10% anger. The main effect of the factor Perceptual history was highly significant ($F(1, 32) = 223.8$, $p < .0001$, $\eta_p^2 = .88$). In the Joy-to-neutral condition, the last expressions were judged as more angry than in the Anger-to-neutral condition. There was no significant main effect of Group ($F(1, 32) = 2.1$, $p = .15$, $\eta_p^2 = .063$). The Endpoint by Group interaction was not significant ($F(2, 64) = .94$, $p = .40$, $\eta_p^2 = .029$), while the Perceptual history by Group interaction was nearly significant ($F(1, 32) = 3.3$, $p = .078$, $\eta_p^2 = .095$). An independent samples t-test revealed that AS individuals rated the last neutral frame of the angry-to-neutral videos as less positive comparing to TD individuals ($t(32) = .81$, $p = .042$). The Endpoint by Perceptual history interaction was significant ($F(2, 64) = 4.1$, $p = .021$, $\eta_p^2 = .11$), reflecting that across both groups the difference between the judgments in the two perceptual histories at the 10% Joy endpoint was larger than at both the neutral endpoint ($t(33) = - 2.7$, $p = .030$) and the 10% anger endpoint ($t(33) = - 2.6$, $p = .013$). Interestingly, the 10% joy endpoint in the joy-video was rated significantly more angry than the 10% anger endpoint in the anger-video ($t(33) = 4.3$, $p < .001$), thus overcoming a 20% difference in intensity. Importantly, the 3-way Endpoint by Perceptual history by Group interaction was non-significant ($F(2, 64) = .063$, $p = .94$, $\eta_p^2 = .002$).

The main conclusion is that the TD and AS groups did not differ in the extent in which their evaluations were influenced by the perceptual histories, at each of the endpoints. This was a surprising finding as the hypothesis was that the AS group would show significantly less overshoot bias than the TD group!

We looked in more detail at the judgments made at the neutral video-endpoints in both groups separately. In the AS group, the mean ratings in the Joy-to-neutral videos (M = 2.73, SD = 0.22) and Anger-to-neutral videos (M = 3.21, SD = 0.28) differed significantly from each other ($t(15) = -6.94$, $p < .0001$), and each of them differed significantly from 3.00 (joy-video: $t(15) = -5.09$, $p < .0001$; anger-video: $t(15) = 2.97$, $p = .010$; two-tailed). The TD group showed a very similar pattern, the mean ratings in the Joy-to-neutral videos (M = 2.71, SD = 0.27) and Anger-to-neutral videos (M = 3.36, SD = 0.31) differed significantly from each other ($t(17) = -7.04$, $p < .0001$), and each of them differed significantly from 3.00 (joy-video: $t(17) = -4.56$, $p < .001$; anger-video: $t(17) = 5.01$, $p < .0001$).

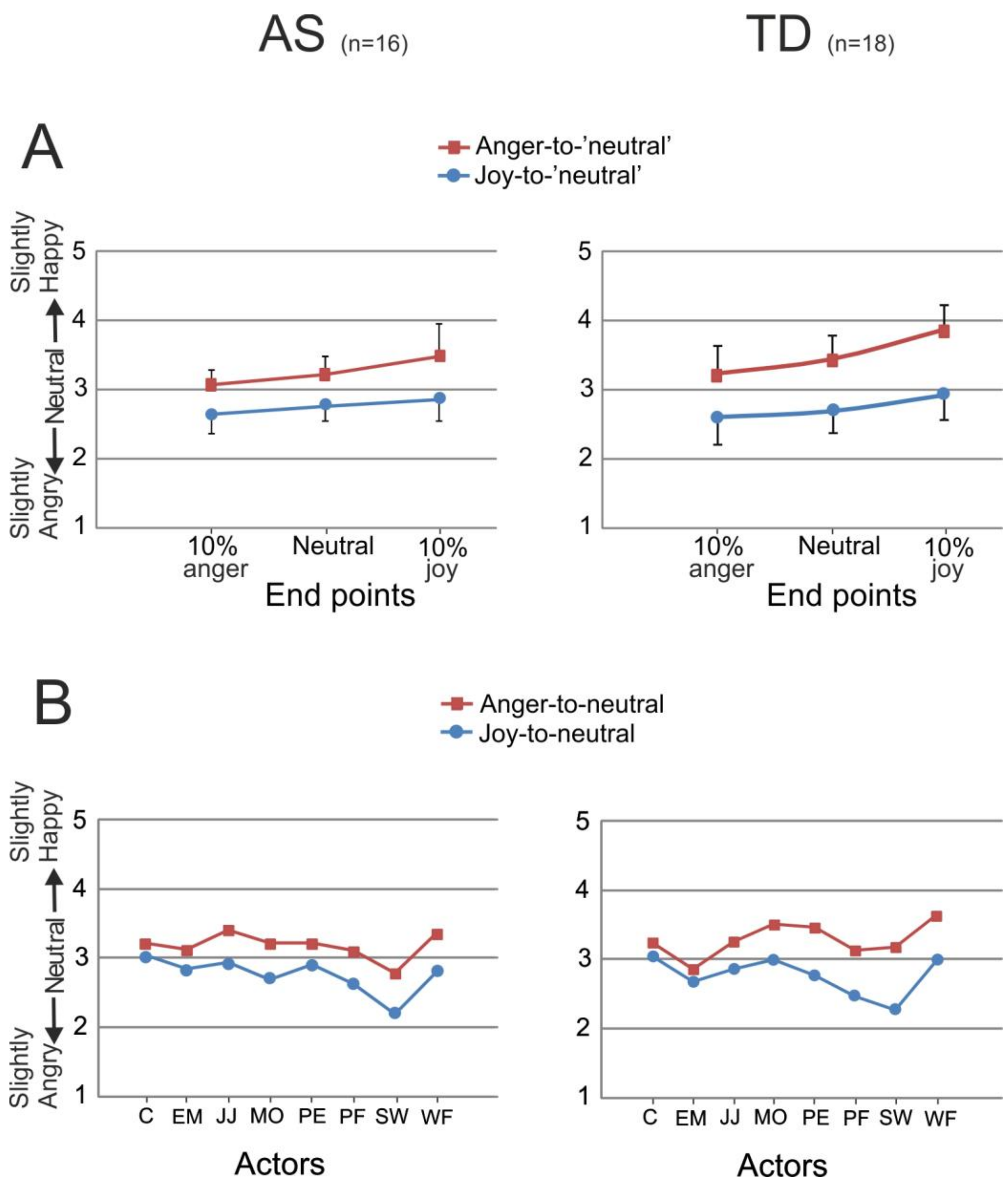

Figure 3.3. Results of Experiment 7. (A) Ratings on the 5-point scale (vertical axis) for the expressions depicted in the last frame of the Joy-to-'neutral' and Anger-to-'neutral' videos for individuals with AS (left side) and for TD individuals (right side). The sequences ended at 10% anger, neutral or 10% joy (horizontal axis). Error bars indicate 1SD. (B) Ratings for exclusively the neutral expressions at the end of the happy- and angry-videos for individuals with AS (left side) and TD individuals (right

side). Scores are shown for each of the eight actors (horizontal axis) to illustrate the consistency.

3.2 3 Discussion

In Experiment 7, both AS and TD participants showed a robust overshoot response bias with the last neutral expression of the Joy-to-neutral videos evaluated as slightly angry and the identical last neutral expression of the Anger-to-neutral videos as slightly happy. Contrary to the initial prediction, the extent of overshoot bias in AS individuals did not differ from that in TD individuals. Further, in both groups the perceptual distortion was strong enough to overcome a 20% difference in emotional intensity, as the 10% joy endpoint in the Joy-video was rated significantly more angry than the 10% anger endpoint in the Anger-video.

If the AS group had shown no overshoot bias, then that would have suggested that they do not engage in an ‘emotional anticipation’ process. However, now that it is clear that they do show an overshoot bias, no firm conclusion can yet be drawn. It could be that they do engage in involuntary emotional anticipation, or that they do not, but use a compensatory mechanism resulting in the same overshoot bias as the TD group obtained through emotional anticipation. The results seem to be in line with a recent study by Uono, Sato, & Toichi (2010) who reported that a group of individuals with developmental pervasive disorders overestimated the intensity of the expression in neutral-to-emotion video-clips in the same way as individuals with TD.

Experiments 8, 9 and 10 were conducted to clarify this point. Specifically, one possibility these experiments looked into was that individuals with AS may have processed the dynamic facial expressions only in terms of physical, non-social, characteristics, without keeping track of the changes in the agent's emotional state of mind as reflected by the changes in the facial expression. For example, they may have relied on the contrast between the first maximally intense expression (happy or angry) and the last neutral frame of the video-clip (see Suzuki & Cavanagh, 1998; Tanaka-Matsumi et al., 1995; Thayer, 1980). However, other mechanisms, such as extrapolation of curves or RM (Yoshikawa & Sato, 2008) might also be able to explain an overshoot bias in individuals with AS.

3.3 Experiment 8: Identity change

Experiment 8 was conducted to test whether compensatory mechanisms based on contrast effects, or on RM effects operating on an underlying positive-negative valence dimension, may explain the overshoot effect found in Experiment 7 in individuals with AS.

In Experiment 8 we changed the identity of the actor in the last frame of the clips, similar to the manipulation performed in Experiment 2 (Chapter 2). Thus, participants needed to judge the last expression of the videos depicting a new identity for which no immediate perceptual history was available. If the overshoot bias in AS individuals resulted from sequential contrast effects or from RM operating on an underlying positive-negative valence dimension, rather than from emotional anticipation, then a change in identity should not affect it. On the other hand, if the bias in the AS group did result from emotional anticipation, then they should not show an overshoot response bias, similarly to the TD group in Experiment 2 (Chapter 2). After all, one cannot anticipate the emotional state of someone on the basis of a perceptual history that pertains to a different identity. Finally, in the condition where the agent's identity did not change (Same-identity condition) we expected to replicate the results found in Experiment 7.

3.3.1 Method

3.3.1.1 Participants

As in Experiment 7, all AS and TD participants had normal or corrected-to-normal vision, and provided written consent prior to the experiment. Participants received course credits or a fee for taking part. The study was approved by the Ethics committee of Hull University.

AS group

Twenty individuals with AS took part in Experiment 8. On the basis of the experimental criteria, three participants were excluded from the analysis (see Data reduction below). The remaining group of 17 AS participants had a mean age of 22.1 years (SD = 7.2 years), consisted of 5 females and 12 males; they had a mean total ADOS score of 8.1 (SD = 0.9), a mean AQ score of 29.0 (SD = 8.2), and a mean total IQ score of 115.9 (SD = 8.1).

TD group

The control group consisted of twenty-one undergraduate Psychology students at Hull University. No one of these participants took part in any of the other experiments. The TD group was recruited to match the AS group for age and gender. After applying exclusion criteria (see Data reduction below), seventeen participants were included in the experiment (age, M = 22.1 years, SD = 5.5 years; 6

females, 11 males; with an AQ score of 15.5 (*SD* = 6.6) and a mean total IQ score of 114.5 (*SD* = 7.7); see Table 3.3.

The AS (n = 17) and TD (n=17) groups did not differ in terms of age ($t(32) = -.675$, $p = .10$), gender ratio ($X^2(1, 34) = .134$, $p = .714$) or IQ ($t(32) = -.54$, $p = .591$). The AQ scores were significantly higher in the AS group than in the TD group ($t(32) = -5.28$, $p < .0001$).

Table 3.3. AS and TD groups in Expriment 8: participant characteristics. Age is in years. Standard deviations are shown between brackets. F, female; M, male; AQ, Autism-spectrum Quotient; IQ-T, total IQ score; IQ-V, verbal IQ score; IQ-P, Performance IQ score; ADOS, autism diagnostic observation schedule.

Group	n	Age	Sex	AQ	IQ-TOT	IQ-V	IQ-P	ADOS
AS	17	22.1 (7.2)	5 F; 12 M	29.0 (8.2)	115.9 (8.1)	117.2 (13.2)	112.1 (15.0)	8.1 (0.9)
TD	17	22.1 (5.5)	6 F; 11 M	15.5 (6.6)	114.5 (7.7)	116.6 (6.4)	112.6 (12.6)	

3.3.1.2 Stimuli and experimental procedure

The stimuli and experimental procedure were similar to those used in Experiment 2 (Chapter 2). The control condition (Same-identity) involved the same videos ending with a neutral point as used in Experiment 7. In the Identity-change condition, the identity of the actor changed in the last frame of the sequence, similarly to the ‘Instant identity change’ condition of Experiment 2. The change in identity was instantaneous, i.e. without inserting morphs in between the two identities to make the transition smooth. An illustration of the stimuli is given in Figure 3.4.

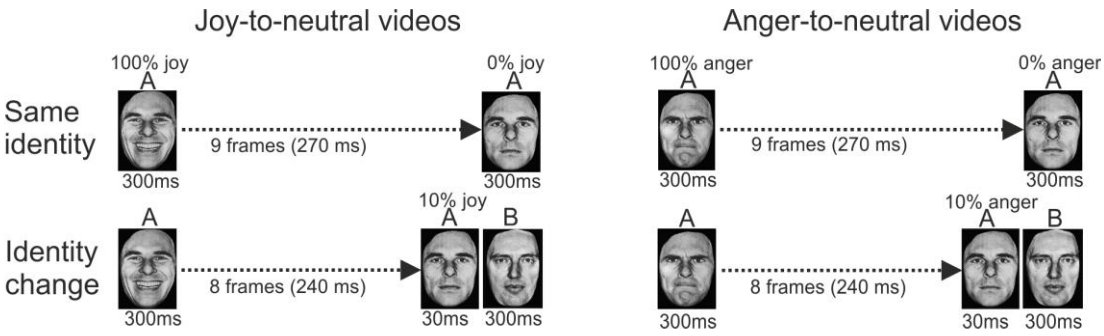

Figure 3.4. Schematic illustration of the stimulus presentations in Experiment 8. Shown are the Same-identity condition (top), and the Identity-change condition (bottom), for Joy-to-neutral (left side) and Anger-to-neutral (right side) perceptual histories.

3.3.1.3 Experimental design

All Participants first rated the neutral expression of each actor presented in isolation as in Experiment 7 (16 trials). Then, following 8 practice trials, 64 experimental trials for each condition were presented in random order (8 actors x 2 perceptual histories x 2 conditions x 4 repetitions). However, in this particular experiment, the AS participants did not perform the control condition (Same-identity), as they had performed this condition in either Experiment 7 or Experiment 9.

3.3.1.4 Data reduction and analysis

Exclusion criteria were the same as for Experiment 7. In the AS group, trials in which the RTs fell below 250 ms or above 3000 ms were removed (2.3% of total number of trials). Participants were excluded if more than 25% of the RTs values were outside the above range (n = 0) and if they exceeded the $\pm$2.5 SD rule to the mean difference ratings (n = 2). Two participants were removed as they pressed the key corresponding to 'neutral' on the scale (3 = neutral) on more than 90% of total number of trials per condition.

In the TD group the RTs criteria excluded 5.7% of the total trials. One participant was removed as more than 25% of the RTs values were outside the stated range. The $\pm$2.5 SD rule to the mean difference ratings excluded other 2 participants. All TD participants used the key button appropriately to express judgments.

In both the groups the calibration scores were used to adjust the scores in the experimental trials; a calibration factor [3.00 minus the calibration score] was added to the experimental scores. All statistical analyses were performed on the calibrated scores.

3.3.2 Results

We initially performed an overall 2x2x2 ANOVA with Task (Same-identity vs. Identity-change) and Perceptual history (Joy-to-neutral video vs. Anger-to-neutral video) as within-subject factors, and Group (AS vs. TD) as between-subjects factor.

The main effect of Task was non-significant ($F(1, 32) = 1.02$, $p = .32$, $\eta_p^2 = .031$). The main effect of Perceptual history was highly significant ($F(1, 32) = 65.9$, $p < .0001$, $\eta_p^2 = .67$), reflecting that the last frame of the videos was judged as more 'happy' in the Anger-to-neutral condition than in the Joy-to-neutral condition. The main effect of Group was also significant ($F(1, 32) = 1.05$, $p = .010$, $\eta_p^2 = .19$), with the AS group scoring lower (i.e. more 'anger' judgments) than the TD group.

The Task by Perceptual history interaction factor was significant ($F(1, 32) = 28.9$, $p < .0001$, $\eta_p^2 = .48$), while the other 2 way interaction factors were non significant (Task by Group: $F(1, 32) = 3.8$, $p = .060$, $\eta_p^2 = .106$; Perceptual history by Group: $F(1, 32) = 1.44$, $p = .24$, $\eta_p^2 = .04$. The 3-way interaction factor was significant ($F(1, 32) = 5.4$, $p = .027$, $\eta_p^2 = .14$).

Next, the two tasks were analyzed separately. The results are illustrated in Figure 3.5.

Same-identity task

A 2X2 mixed ANOVA with Perceptual history (Joy-to-neutral video vs. Anger-to-neutral video) as the within-subject factor, and Group (AS vs. TD) as between-subjects factor showed a significant main effect of Perceptual history ($F(1, 32) = 70.4$, $p < .0001$, $\eta_p^2 = .68$), reflecting that the last frame of the videos was judged as more 'happy' in the Anger-to-neutral condition than in the Joy-to-neutral condition. A main effect of Group was not significant ($F(1, 32) = 3.23$, $p = .082$, $\eta_p^2 = .092$). The Perceptual history by Group interaction was not significant ($F(1, 32) = .25$, $p = .62$, $\eta_p^2 = .008$). Thus, on the Same-identity task the TD and AS groups showed an almost identical performance, both showing a robust overshoot response bias.

Identity-change task

The 2X2 mixed ANOVA showed significant main effects of both Perceptual history ($F(1, 32) = 11.17$, $p = .002$, $\eta_p^2 = .259$) and Group $F(1, 32) = 10.02$, $p = .003$, $\eta_p^2 = .239$). Crucially, the Perceptual history by Group interaction was significant ($F(1, 32) = 8.70$, $p = .006$, $\eta_p^2 = .214$).

In the TD group, the ratings for the neutral expressions did not differ between the two perceptual histories ($t(16) = -.437$, $p = .668$). Thus, in TD participants a change of identity of the actor effectively removed the overshoot bias. In the AS group, however, the ratings in the Anger-to-neutral videos were significantly higher (i.e. more 'happy') than in the Joy-to-neutral videos ($t(16) = 3.52$, $p = .003$).

Thus, while in the Same-identity condition the two groups performed almost identically, in the Identity-change condition the groups performed significantly different, with the AS group, but not the TD group, showing an overshoot response bias.

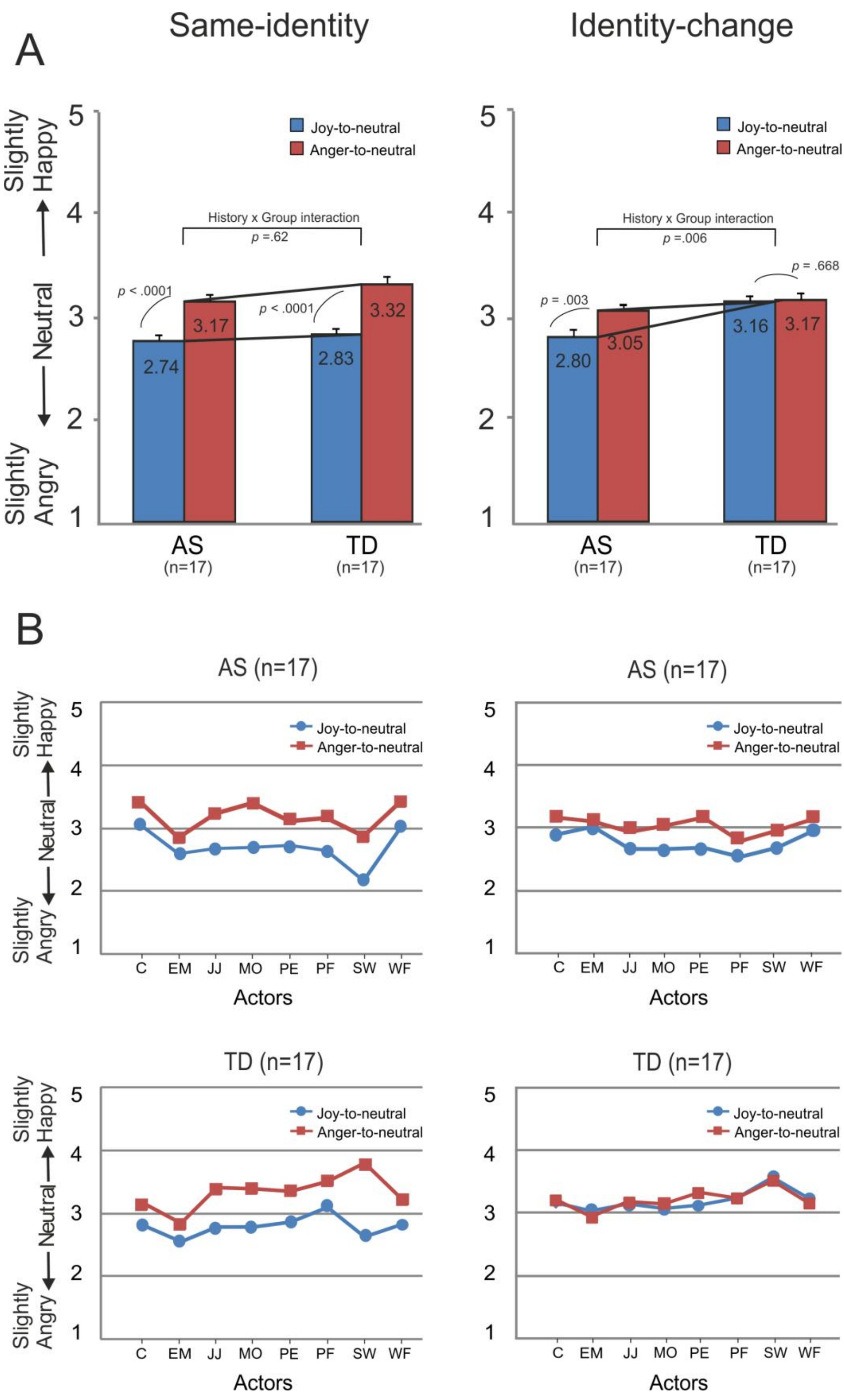
A
Same-identity
Identity-change
Slightly Happy
Neutral
Slightly Angry
Joy-to-neutral
Anger-to-neutral
History x Group interaction
p =.62
p =.006
p < .0001
p < .0001
p = .003
p = .668
2.74
3.17
2.83
3.32
2.80
3.05
3.16
3.17
AS (n=17)
TD (n=17)
B
AS (n=17)
TD (n=17)
C EM JJ MO PE PF SW WF
Actors

Figure 3.5. Results of Experiment 8. (A) Scores for the neutral expressions at the end of the Joy-to-neutral and Anger-to-neutral sequences in the same-identity (left side) and identity-change (right side) conditions with AS and TD individuals (horizontal axis). Error bars indicate SEM (standard error of the mean). (B) Ratings for each of the eight actors in the same-identity (left side) and identity-change (right side) conditions are shown to illustrate the consistency across actors with AS (left panel) and TD (right panel) individuals.

Could poor identity recognition skills in the AS group explain the presence of an overshoot bias in this group in the Identity-change condition?

In principle it might be possible that the overshoot bias found in the Identity-change condition in AS individuals could result from impaired identity recognition skills (cf. Behrmann, Thomas, & Humphreys, 2006). If the AS participants had not detected the identity change in some of the trials, then they would have treated these trials as if they were Same-identity trials. In the Same-identity condition they produce a robust overshoot bias, similarly to the TD group (Figure 3.5, left panel). We reasoned that if the overshoot bias in the Identity-change condition was due to a difficulty of AS participants in recognizing the difference between the two identities, then more overshoot was expected in trials where the two identities were both females or both males, as compared to trials in which there was a gender shift (M>F, F>M). Recent research on the gender recognition issue in autism shed some light on whether these individuals are impaired in discriminating faces of different gender. For instance, using a gender recognition task, Best, Minshew, & Strauss (2010) found that adults with HF autism can detect gender from the eyes, although with less success as compared to TD adults.

Moreover, it has been reported that individuals with autism perform worse at discriminating gender from the whole face only when faces lack typical gender information (Newell, Best, Gastgeb, Rump, & Strauss, 2011).

Therefore, the idea in the current experiment was that a change in gender would have alerted the participant to the change in identity. Furthermore, in terms of shapes and physical features, a male and a female face would look more dissimilar than two male, or two female, faces.

When there was no switch in gender between the two identities (M>M, F>F), the mean score of the AS group in the Joy-to-neutral videos was 2.77 (SD = 0.25), and the mean score in the Anger-to-neutral videos 3.01 (SD = 0.20), which did differ significantly from each other (t(16) = -3.85, *p* = .003). When there was a switch in gender between the two identities (M>F, F>M), the mean score in the Joy-to-neutral videos was 2.88 (SD = 0.32), and the mean score in the Anger-to-neutral videos was 3.05 (SD = 0.29), which again differed significantly from each other (t(16) = -3.33, *p* = .004) (see Figure 3.6).

Comparing the mean difference scores [i.e. the mean score in Anger-to-neutral videos minus the mean score in Joy-to-neutral videos] in the trials with and without a gender switch showed that these did not differ from each other (t(16) = 1.23, *p* = .237).

This may suggest that poor identity recognition skills did not play a role in bringing about the overshoot bias in the Identity-change condition in AS participants.

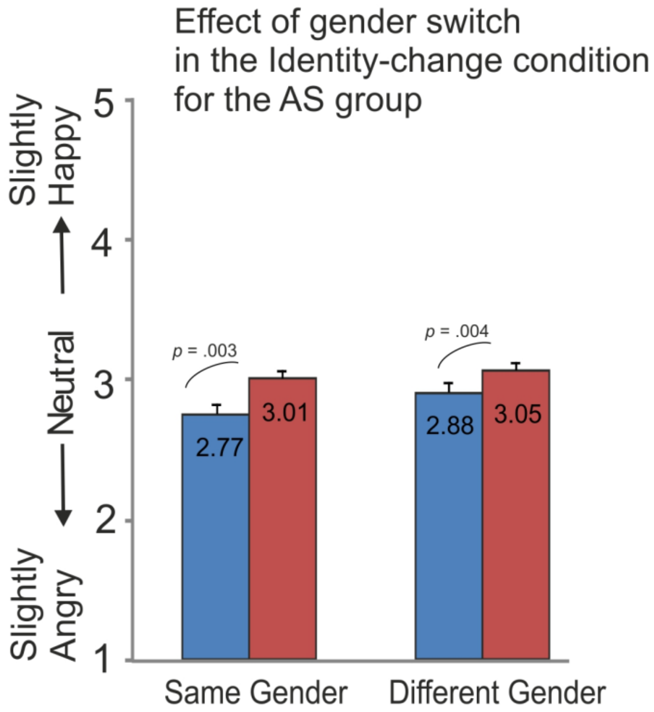

Figure 3.6. Results of the gender switch effect in Experiment 8. Scores for the neutral expressions at the end of the Joy-to-neutral and Anger-to-neutral sequences are shown in the identity-change condition with same gender-change and different gender-change (horizontal axis) in AS individuals.

3.3.3 Discussion

When the video-clips depicted the same identity throughout, both AS and TD individuals produced a robust overshoot response bias, confirming the results of Experiment 7. Interestingly, following the Identity-change manipulation, the overshoot bias was still present in the AS group, but was completely absent in the TD group. We further demonstrated that the overshoot response bias found in AS participants in the Identity-change condition cannot be explained by a difficulty in recognizing different identities.

On themselves, the results of the Same-identity condition might suggest that AS individuals show emotional anticipation, similarly to TD individuals. However, the results of the Identity-change condition show that this is probably not the case. The AS group continued to produce an overshoot bias in the Identity-change condition, whereas the TD group did not. Emotional anticipation would not produce an overshoot bias in response to agent B about whom the participant knows nothing, other than that agent B depicts a neutral expression. The perceptual history pertained exclusively to agent A, but no questions are asked about agent A, the evaluation has to be made about agent's B expression. These findings open up the possibility that the AS group used an alternative strategy to produce the bias in the Same-identity condition, and simply applied the same alternative strategy in the Identity-change condition, hence their bias in the latter condition.

What might the alternative strategy used by the AS participants be?

One possibility is that the AS group might have relied on the *contrast* in facial expression and emotional valence between the 100% emotional expression shown at the start of the clip (for 300 ms) and the neutral expression shown at the end of the clip (also presented for 300 ms). Another strategy might involve an extrapolation of the dynamic geometries in the face, or an extrapolation on an underlying positive-negative valence dimension. These two strategies involve representational momentum (RM; Freyd, 1987; Freyd & Finke, 1984). The underlying positive-negative valence dimension, and to a large extent also the extrapolation of the dynamic geometries from the face, remained unaffected by the identity change.

The absence of an overshoot bias in TD individuals in the Identity-change condition confirmed the findings reported in Experiment 2 (Chapter 2). It suggests that the large overshoot in the Same-identity condition was not produced by sequential contrast effects or RM. Because, if TD participants relied on these low-level bottom-up processes, then there is no reason why they should not also be affected by it in the Identity-change condition. This finding favours a top-down emotional anticipation explanation.

3.4 Experiment 9: Mask

Experiment 9 aimed to test whether the overshoot response bias found in AS individuals, was related to the uninterrupted flow of motion that extended up to the last, neutral expression. Such a flow of motion might have facilitated an extrapolation of facial geometries (cf. Enns et al., 2010). In the face-sequences presented, many geometrical features changed in a continuous and smooth manner. For example, the U-shaped mouth in the smile gradually transformed into a flat-shaped mouth in the neutral expression, which might be extrapolated into a slightly inverted U shape, giving the appearance of a slightly angry expression. In Experiment 9 the continuous flow in the video-sequence was disrupted by inserting a 400 ms mask immediately before this last neutral frame (Mask condition). If such a perceptual mechanism played a crucial role in bringing about the bias in AS individuals, then one would expect not to find an overshoot bias in the Mask condition in this group. For the TD group, the interruption should not affect the anticipation of the emotional state of the actor, which is based on the content present in the immediate perceptual history. The mask would then only act as a temporary occluder. Therefore the emotional anticipation hypothesis predicts to find an overshoot bias.

3.4.1 Method

3.4.1.1 Participants

As in Experiments 7 and 8, all AS and TD participants had normal or corrected-to-normal vision, and provided written consent prior to the experiment. Participants received course credits or a fee for taking part. The study was approved by the Ethics committee of Hull University.

AS group

Fifteen AS individuals were tested for Experiment 9. All these AS participants did both the Mask and No-Mask conditions (see below). On the basis of the experimental criteria (see Data reduction below) one participant was excluded from the analysis. The fourteen participants included (4 females, 10 males; mean age = 22.2 years, SD = 4.7) had a mean total ADOS score of 8.0 (SD = 0.9), a mean IQ score of 118.3 (SD = 6.9), and a mean AQ score of 30.8 (SD = 7.4).

TD group

Sixteen undergraduate Psychology students at Hull University took part. None of these participants took part in any of the other experiments. The TD group was matched to the AS group for age, gender and IQ. After applying exclusion criteria, fourteen participants were included in the analysis (age, M = 20.3 years, SD = 3.1; 5

females, 9 males; with an AQ score of 15.5 (*SD* = 4.9) and a mean total IQ score of 113.7 (*SD* = 7.6); see Table 3.4.

The TD and AS groups did not differ in terms of age ($t(26) = -1.23$, $p = .272$), gender ratio ($X^2(1, 28) = .164$, $p = .686$) nor IQ ($t(26) = -1.67$, $p = .106$). The AQ scores were significantly higher in the AS group than in the TD group ($t(26) = -6.43$, $p < .0001$).

Table 3.4. AS and TD groups in Expriment 9: participant characteristics. Age is in years. Standard deviations are shown between brackets. F, female; M, male; AQ, Autism-spectrum Quotient; IQ-T, total IQ score; IQ-V, verbal IQ score; IQ-P, Performance IQ score; ADOS, autism diagnostic observation schedule.

Group	n	Age	Sex	AQ	IQ-TOT	IQ-V	IQ-P	ADOS
AS	14	22.2 (4.7)	4 F; 10 M	30.8 (7.4)	118.3 (6.9)	118.7 (11.3)	114.3 (10.7)	8.0 (0.9)
TD	14	20.3 (3.1)	5 F; 9 M	15.5 (4.9)	113.7 (7.6)	114.3 (8.2)	111.9 (10.4)	

3.4.1.2 Stimuli and experimental procedure

The stimuli and experimental procedure were the same as in Experiment 5 (Chapter 2). In short, Joy-to-neutral and Anger-to-neutral video-clips were presented. The last neutral frame was immediately preceded by a 400 ms long mask, consisting of an oval grid in grey scale. The total duration of the sequence, mask included, was 1270 ms. The control condition was identical to the mask condition except that no mask was inserted. An illustration of the video-sequences is given in Figure 3.7.

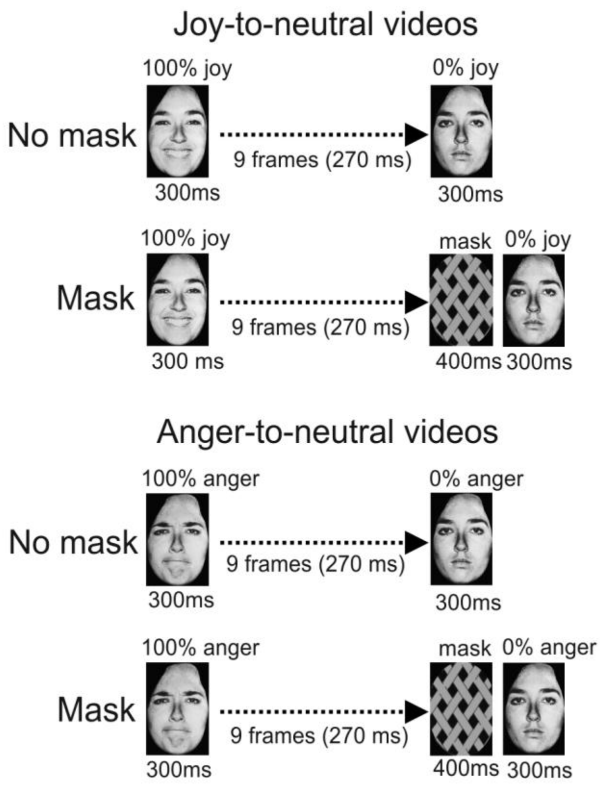

Figure 3.7. Schematic illustration of the stimuli presentations in Experiment 9. Joy-to-neutral videos (top panel) and Anger-to-neutral videos (bottom panel) in the No-Mask (Control) and Mask conditions are shown.

3.4.1.3 Experimental design

Participants first rated the neutral expression of each actor presented in isolation (16 trials) as in the previous experiments. A 2x2x2 factorial design was used with Condition (No mask vs. Mask) and Perceptual history (Joy-to-neutral vs. Anger-to-neutral) and Group (AS vs. TD) as factors. The experimental block consisted of 32 trials for both the No-Mask and Mask conditions (8 actors x 2 perceptual histories x 2 repetitions).

3.4.1.4 Data reduction and analysis

Exclusion criteria were the same as for Experiments 7 and 8. In the AS group, trials in which the RTs fell below 250 ms or above 3000 ms were removed (4.5% of total number of trials). Participants were excluded if more than 25% of the RTs values were outside this range (n = 0). A ±2.5 SD rule to the mean difference ratings excluded 1 participant.

In the TD group, 8.9% of the total number of trials fell below 250 ms or above 3000 ms, and was removed. Participants were excluded if more than 25% of the RTs values were outside the above criteria (n = 0). The ±2.5 SD rule to the mean difference ratings excluded 2 participants. No anomalies were recorded in the way both AS and TD participants selected the keys to express their judgments.

In both groups the calibration scores were used to adjust the scores in the experimental trials; a calibration factor [3.00 minus the calibration score] was added to the experimental scores. All statistical analyses were performed on the calibrated scores.

3.4.2 Results

The results are illustrated in Figure 3.8. A 2X2x2 mixed ANOVA with Task (No-Mask vs. Mask) and Perceptual history (Joy-to-neutral video vs. Anger-to-neutral video) as the within-subject factors, and Group (TD vs. AS) as between-subjects factor, revealed no significant main effect of Task ($F(1, 26) = .257$, $p = .62$, $\eta_p^2 = .010$). The main effect of Perceptual history was highly significant ($F(1, 26) = 151.2$, $p < .0001$, $\eta_p^2 = .85$); the main effect of the between-subjects factor Group was not significant ($F(1, 26) = .78$, $p = .39$, $\eta_p^2 = .029$). Of the two-way interaction factors, the

Perceptual history by Group interaction was significant ($F(1, 26) = 10.9$, $p = .003$, $\eta_p^2 = .29$). This indicated that, across both tasks, the overshoot response bias (i.e. score in the Joy-to-neutral condition minus score in the Anger-to-neutral condition) was larger in the TD group than in the AS group ($t(26) = 3.30$, $p = .003$). The other two-way interaction factors were non-significant (Task by Group: ($F(1, 26) = 1.1$, $p = .30$, $\eta_p^2 = .041$; Task by Perceptual history: $F(1, 26) = .031$, $p = .86$, $\eta_p^2 = .001$). The 3-way Task by Perceptual history by Group interaction was also non-significant ($F(1, 26) = 1.7$, $p = .20$, $\eta_p^2 = .063$). Thus, The AS and TD groups performed very similarly in this experiment.

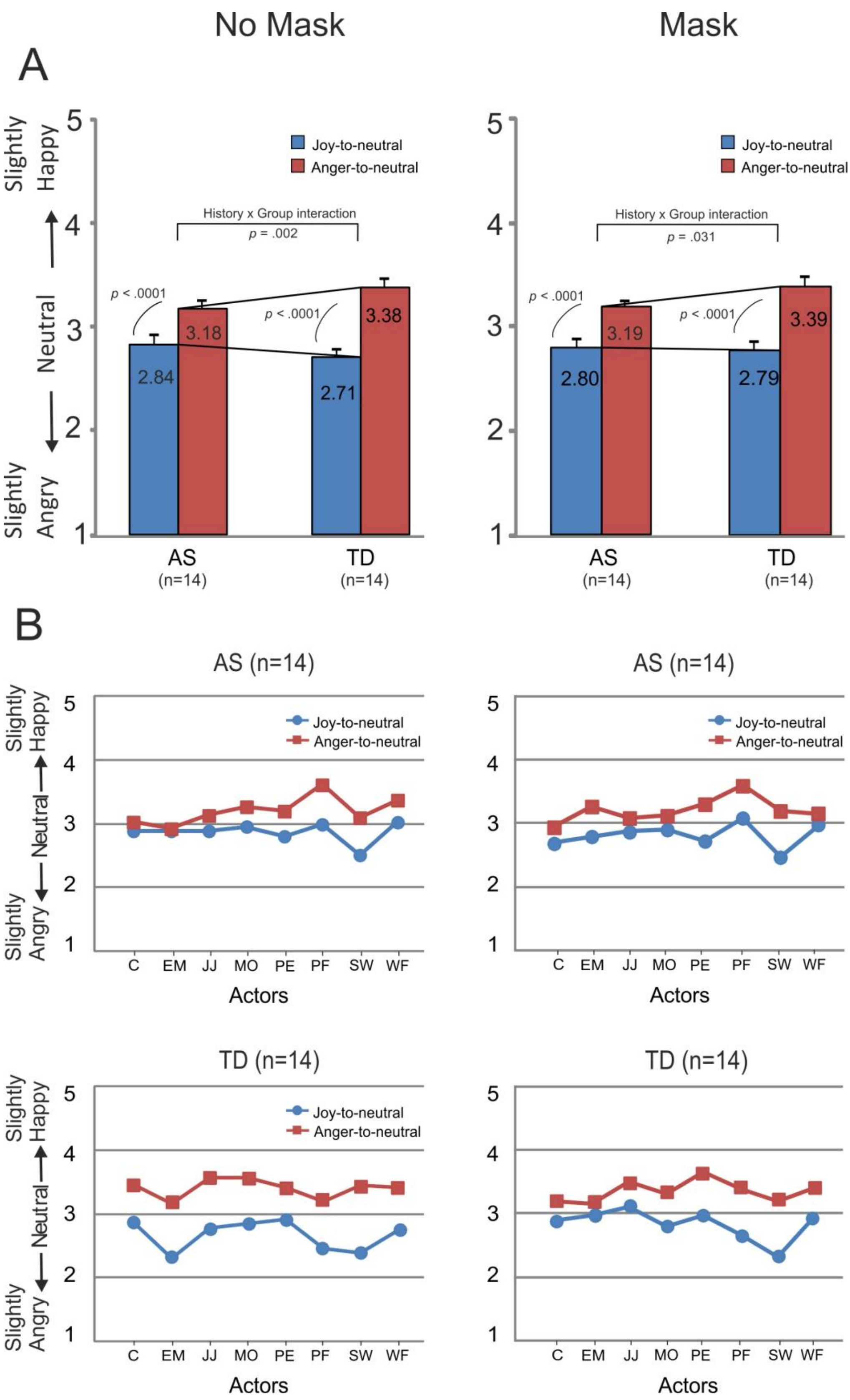
No Mask
Mask
A
Slightly Happy
Neutral
Slightly Angry
Joy-to-neutral
Anger-to-neutral
History x Group interaction
p = .002
p = .031
p < .0001
2.84
3.18
2.71
3.38
2.80
3.19
2.79
3.39
AS (n=14)
TD (n=14)
B
AS (n=14)
TD (n=14)
C EM JJ MO PE PF SW WF
Actors

Figure 3.8. Results of Experiment 9. (A) Scores for the neutral expressions at the end of the Joy-to-neutral and Anger-to-neutral sequences in the no-mask (left side) and mask (right side) conditions with AS and TD individuals (horizontal axis). Error bars indicate SEM (standard error of the mean). (B) Ratings for each of the eight actors in the no-mask (left side) and mask (right side) conditions are shown to illustrate the consistency across actors with AS (top panel) and TD (bottom panel) individuals.

3.4.3 Discussion

Experiment 9 showed that the overshoot bias did survive the insertion of the mask in both groups. As the mask manipulation should have reset low-level mechanisms that rely on the flow of motion (cf. Enns et al., 2010), we conclude that an extrapolation of curves present in the face cannot explain the overshoot bias in both participant groups. The results of Experiment 9 for the TD participants replicate previous findings (Experiment 5, Chapter 2) and are congruent with the emotional anticipation hypothesis.

In the control condition (No Mask) of the current experiment, the response bias appeared less robust in the AS group compared to the TD group. This contrasts somewhat with the findings in Experiments 7 and 8, in which the magnitude of the overshoot response bias in the TD and AS groups was quite similar. However, the performance of the AS group was quite consistent across these experiments. The group of AS individuals that did the control condition in Experiment 7 (n=16) and the group of AS individuals that did the control condition in Experiment 8 (n=14) largely overlapped. Of the 14 AS participants in Experiment 9, four participants were new and had not been tested in Experiment 7 (there was a 10 month time interval between the conduction of Experiments 7 and 9). The overshoot biases of these two AS groups did not differ from each other (t(28) = 1.4, *p* = .16, independent sample t test, two-tailed). The TD control groups used in Experiments 7, 8 and 9 did, however, not overlap at all (none of the TD participants was tested in more than one experiment). Therefore, the slight discrepancy between Experiments 7, 8 and 9 was due to variability in the TD control groups.

Experiment 10 was conducted to specifically clarify the role played by contrast effects in bringing about the overshoot bias in the AS group.

3.5 Experiment 10: Loop

In Experiment 10 the video-sequence was modified so that it started from a neutral expression and morphed via a maximally happy or angry expression back to the same neutral expression (Loop condition). The rational was that if the response bias found in the previous experiments in AS individuals was due to the contrast between the - to be evaluated - last neutral expression and the ‘anchor’ stimulus (the first happy or angry expression presented for 300 ms), then in the loop condition one would expect this last neutral expression to be judged as neutral and not obtain an overshoot bias, even though the second half of the sequence – Joy/Anger-to-neutral - is identical to the sequence in the previous experiments.

Our predictions for TD individuals were based on the results found in Experiment 3 (Chapter 2), where TD participants showed an ‘undershoot’, effect (i.e. a bias which was opposite in direction to the usual overshoot).

3.5.1 Method

3.5.1.1 Participants

As in the previous experiments, all AS and TD participants had normal or corrected-to-normal vision, and provided written consent prior to the experiment. Participants received course credits or a fee for taking part. The study was approved by the Ethics committee of Hull University.

AS group

Eighteen AS individuals out of the group recruited for the previous experiments made Experiment 10. As for Experiment 2 (Chapter 3), considering various difficulties related to the recruitment, these participants made only the experimental condition (Loop) of Experiment 10. On the basis of the experimental criteria (see Data reduction below) three students were excluded from the analysis. Fifteen AS individuals were included (5 females, 10 males; mean age = 22.3 years, SD = 7.1). They had a mean total ADOS score of 8.2 (SD = 1.6), a mean IQ score of 118.9 (SD = 8.9), and a mean AQ score of 33.7 (SD = 5.7).

TD group

Twenty-two undergraduate Psychology students at Hull University took part in the experiment. No one of these participants took part in any of other experiments. The TD group matched the AS group for age, gender and IQ (see Table 3.5 for sub-scores). After applying exclusion criteria (see Data reduction below), nineteen participants were included in the analysis (age, M = 20.6 years, SD = 2.52; 6 females, 13 males; with an AQ score of 14.7 (*SD* = 6.1) and a mean total IQ score of 115.1 (*SD* = 8.1).

The TD and AS groups did not differ in terms of age ($t(32) = -1.00$, $p = .323$), gender ratio ($X^2(1, 34) = .012$, $p = .914$) or IQ ($t(32) = -1.49$, $p = .147$). The AQ scores were significantly higher in the AS group than in the TD group ($t(32) =-9.29$, $p < .0001$).

Table 3.5. AS and TD groups in Expriment 10: participant characteristics. Age is in years. Standard deviations are shown between brackets. F, female; M, male; AQ, Autism-spectrum Quotient; IQ-T, total IQ score; IQ-V, verbal IQ score; IQ-P, Performance IQ score; ADOS, autism diagnostic observation schedule.

Group	n	Age	Sex	AQ	IQ-TOT	IQ-V	IQ-P	ADOS
AS	15	22.3 (7.1)	5 F; 10 M	33.7 (5.7)	118.9 (8.9)	121.6 (13.6)	112.3 (11.9)	8.2 (1.6)
TD	19	20.6 (2.5)	6 F; 13 M	14.7 (6.1)	114.7 (7.5)	115.5 (7.4)	112.5 (10.8)	

3.5.1.2 Stimuli and experimental procedure

Stimuli and procedure were identical to those in Experiment 3 (Chapter 2). Video-clips displayed a neutral expression, which morphed via a maximally happy (or angry) expression back to the same neutral expression (Neutral-to-joy-to-neutral and Neutral-to-anger-to-neutral sequences). The morphing sequence consisted of 19 interpolated frames, each 30 ms long. The first and the last frames both lasted 300 ms, making the entire sequence last for 1170 ms. The control condition consisted of Joy-to-neutral and Anger-to-neutral video sequences (see Figure 3.9).

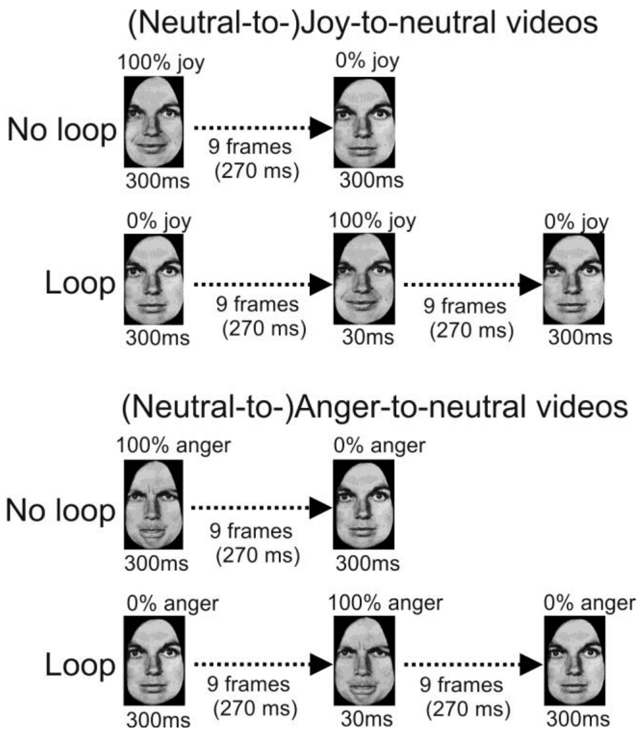

Figure 3.9. Schematic illustration of the stimulus presentations in Experiment 10. (Top panel) Joy-to-neutral sequence in the control condition (top row) and Neutral-to-joy-to-neutral sequence in the loop condition (bottom row). (Bottom panel) Similar sequences but for Anger-to-neutral (top row) and Neutral-to-anger-to-neutral (bottom row).

3.5.1.3 Experimental design

Participants first rated the neutral expression of each actor presented in isolation (16 trials) as in the previous experiments. The experimental design was as in Experiment 7, except that Experiment 10 consisted of 96 randomized trials: 8 actors x 2 perceptual histories x 2 conditions x 3 repetitions. Scores for the AS individuals in the control condition (No Loop) were obtained from either Experiment 7 or Experiment 9.

3.5.1.4 Data reduction and analysis

Exclusion criteria were the same as for previous experiments. In the AS group, in 9.2% of the total number of trials the RTs fell below 250 ms or above 3000 ms and were removed. Participants were excluded if more than 25% of the RTs values were outside this range (n = 0). The $\pm$2.5 SD rule to the mean difference ratings excluded two participants. One more participant was removed because the judgments on the scale were made pressing the same key for more than 90% of total number of trials.

In the TD group, 7.6% of the total number of trials was removed for having RTs outside the above range. Participants were excluded if more than 25% of the RTs values were outside the RTs criteria (n = 2). A $\pm$2.5 SD rule to the mean difference ratings excluded one more participant. No anomalies were recorded in the way as TD participants used the key box for their judgments.

In both the groups the calibration scores were used to adjust the scores in the experimental trials; a calibration factor [3.00 minus the calibration score] was added to the experimental scores. All statistical analyses were performed on the calibrated scores.

3.5.2 Results

First a 2X2x2 mixed ANOVA was performed with Task (No-Loop vs. Loop) and Perceptual history (Joy-to-neutral video vs. Anger-to-neutral video) as within-subject factors, and Group (TD vs. AS) as between-subjects factor. The main effect of Task was non-significant ($F(1, 32) = .038$, $p = .85$, $\eta_p^2 = .001$). The main effect of Perceptual history was also non-significant ($F(1, 32) = .042$, $p = .84$, $\eta_p^2 = .001$). This was quite surprising as so far the main effect of Perceptual history had always been highly significant. It reflected that in the Loop task, the pattern of scores was effectively the opposite of that in the No-Loop task. The main effect of the between-subjects factor Group was not significant ($F(1, 32) = 1as.4$, $p = .24$, $\eta_p^2 = .042$).

Of the two-way interaction factors, the Task by Perceptual history factor was significant ($F(1, 32) = 35.1$, $p < .0001$, $\eta_p^2 = .52$), reflecting the reversal of the pattern between the two tasks, independent of the groups. The other two-way interaction factors were non-significant (Task by Group: ($F(1, 32) = .932$, $p = .34$, $\eta_p^2 = .028$; Perceptual history by Group: $F(1, 32) = 2.08$, $p = .16$, $\eta_p^2 = .061$).

The 3-way interaction of Task by Perceptual history by Group interaction was also non-significant ($F(1, 32) = 1.23$, $p = .28$, $\eta_p^2 = .037$). The results are illustrated in Figure 3.10.

Thus, the AS and TD groups both showed a robust overshoot response bias in the No-Loop task and a respose bias opposite in direction to the overhoot (undershoot) in the Loop task.

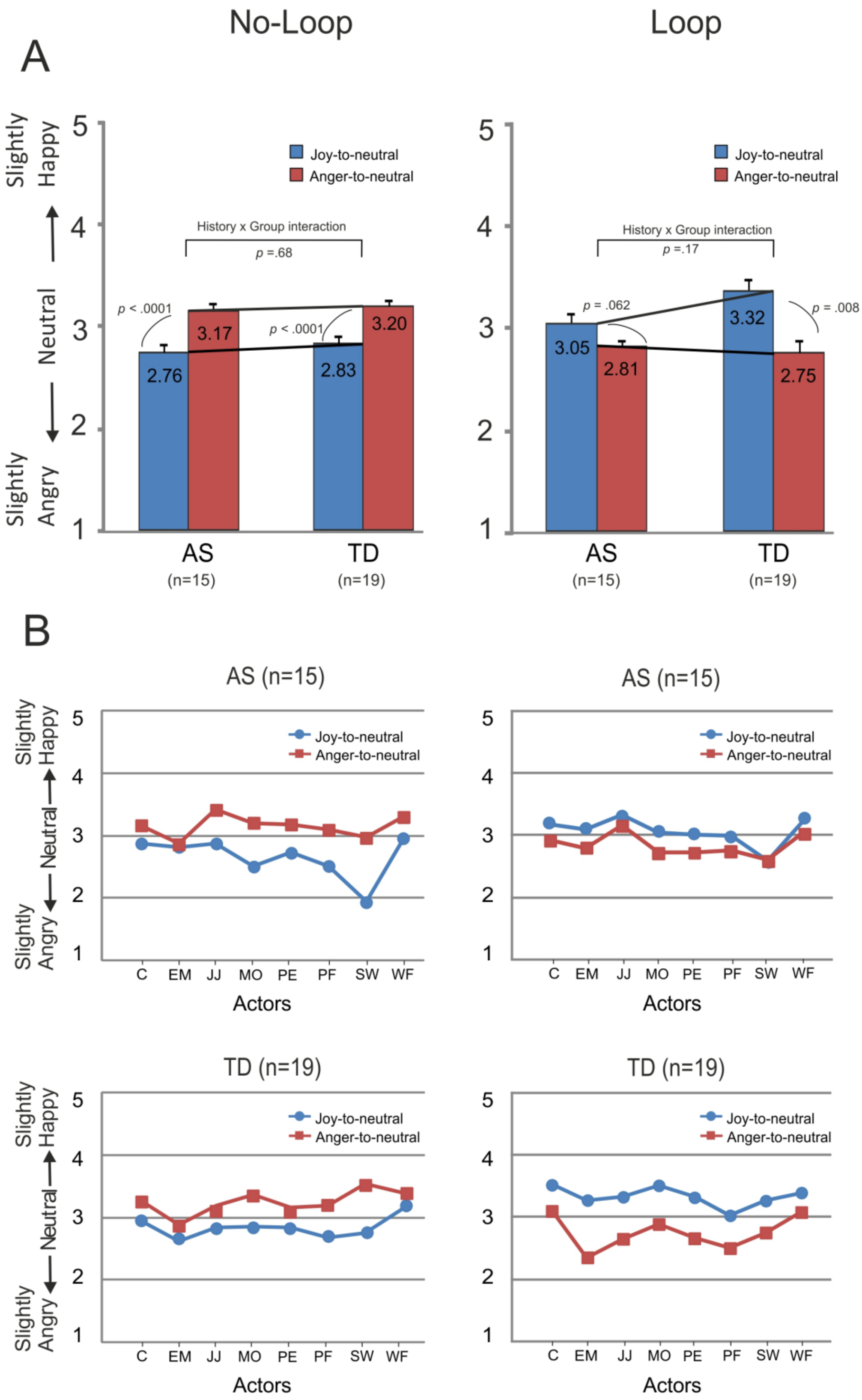
No-Loop
Loop
A
Slightly Happy
Neutral
Slightly Angry
Joy-to-neutral
Anger-to-neutral
History x Group interaction
p =.68
p < .0001
p < .0001
2.76
3.17
2.83
3.20
AS
(n=15)
TD
(n=19)
History x Group interaction
p =.17
p = .062
p = .008
3.05
2.81
3.32
2.75
AS
(n=15)
TD
(n=19)
B
AS (n=15)
AS (n=15)
TD (n=19)
TD (n=19)
C EM JJ MO PE PF SW WF
Actors

Figure 3.10. Results of Experiment 10. (A) Scores for the neutral expressions at the end of the Joy-to-neutral and Anger-to-neutral sequences in the no-loop (left side) and loop (right side) conditions with AS and TD individuals (horizontal axis). Error bars indicate SEM (standard error of the mean). (B) Ratings for each of the eight actors in the no-loop (left side) and loop (right side) conditions are shown to illustrate the consistency across actors with AS (left panel) and TD (right panel) individuals.

3.5.3 Discussion

In Experiment 10 the Loop condition was introduced, which was characterized by the absence of a contrast between the initial frame (300 ms duration) and last frame (300 ms duration) of the video-sequences (both showing identical neutral expressions). First, in the No-Loop task, which constituted a control condition, both groups showed a nearly identical overshoot response bias. This confirmed the findings in Experiments 7, 8 and 9, that AS participants do produce an overshoot bias, which is comparable in size to that of TD participants.

In the Loop task, the overshoot bias had reversed into an ‘undershoot’ bias in both TD and AS groups. However, a closer look at the mean ratings in the AS group for the Neutral-to-joy-to-neutral condition: 3.05 (SD = 0.35), and for the Neutral-to-anger-to-neutral condition: 2.81 (SD = 0.31), revealed that this ‘undersoot’ bias was very subtle ($t(14) = 2.03$, p =.062). On the other hand, in the TD group we did find a strong ‘undershoot’ effect ($t(18) = 2.97$, $p = .008$, $\alpha = .025$). Therefore, in the AS group there was just a tendency towards an undershoot bias.

The current experiment was designed to shed light on the merits of two distinct low-level visual explanations for the overshoot response bias, namely (i) extrapolation or RM, and (ii) sequential contrast effects.

(i)_Extrapolation (or RM) would predict to find an overshoot response bias in the Loop condition, i.e. a continuation or extrapolation of the second half of the sequence, which went from maximal joy or anger to a neutral expression. We did not obtain such an overshoot bias, but, in contrast, obtained an 'undershoot'. This strongly suggests that extrapolation (or RM) was not the mechanism underpinning the overshoot bias in the No-Loop condition.

(ii)_The sequential contrast hypothesis predicted that the last frame would be judged as neutral, as there was no contrast between the first and last frames of the video-sequence (both being neutral). This did not happen as there was a clear trend for the last frame to be judged as 'slightly happy' at the end of the Neutral-to-joy-to-neutral sequence and as slightly angry at the end of the Neutral-to-anger-to-neutral sequence. This was found in both groups, although, as shown above, to a lesser extent in the AS group. This result opens up the possibility that the overshoot effect as found in the No loop task with AS individuals may be explained by sequential contrast while it rules out a mayor contribution of contrast effects to the overshoot bias in TD individuals.

How should we interpret the 'undershoot' response bias? One possibility is that the video-sequences produced an either positive or negative 'social signal'. The neutral-joy-neutral sequence might produce a positive social signal directed towards the observer, causing the observer to judge this person 'positively' in return. Similarly, the neutral-anger-neutral sequence would produce a negative social signal towards the observer, causing the observer to judge this person 'negatively'. In this scenario, the observer simply receives knowledge that the agent either likes or dislikes him/her, and, as we tend to like people who like us and dislike those who dislike us (Jones et al., 2006), the favor is returned. However, this does not necessarily mean that the observer shows to be able of ‘emotional anticipation’. The social signal *per se* is enough to influence the evaluations of the agent’s neutral face.

Thus both groups processed these social signals and were influenced by them, although the TD individuals were influenced to a larger extent than the AS individuals.

Alternatively, the AS group could have used an altogether different mechanism to produce the undershoot response bias. Possibly, individuals with AS were more sensitive to the pattern present in the video and expected this pattern to repeat itself. However, the lack of a strong ‘undershoot’ effect favours the sequential contrast hypothesis.

3.6 General discussion

Four different low-level visual mechanisms can be identified, each of which could explain the overshoot bias: (1) Contrast effects, (2) RM on a positive-negative valence dimension, (3) RM on facial features and (4) adaptation. Three experimental manipulations were applied to test the merits of these mechanisms: Identity change (IC, Experiment 8), insertion of Mask (Experiment 9), and the Loop sequence (Experiment 10). In addition, the merits of the emotional anticipation hypothesis were examined. During social interactions we do not perceive faces with fixed expressions, rather we are presented with innumerable different shades of expressions that change over time. The ability to attribute the emotional states of mind underpinning those expressions to the agent expressing them, and to involuntary anticipate how these most likely would change in the immediate future, allows us to engage successfully in social interactions.

In Chapter 2 it was suggested that the dynamic perceptual history of a target neutral face gives rise to an involuntary anticipation of how the emotional state of the agent would change next, which in turn biases the perception of that neutral face (‘overshoot’ response bias). We propose that such an effect is the result of a low-level emotion-reading process (Goldman, 2006; Goldman & Sripada, 2005), rather than the result of basic visual (bottom-up) processes. This emotion reading process is a top-down mechanism, which we called ‘emotional anticipation’,

In Chapter 3 we were interested to see whether individuals with Asperger’s syndrome (AS) use ‘emotional anticipation’ too, or whether they might use alternative mechanisms to solve our task. Their proposed lack of intuitive and involuntary understanding of other people’s behavior (Jellema et al., 2009; Kuchinke et al., 2011; Kuzmanovic et al., 2011; Senju et al., 2009) might suggest they meet the latter option.

The Typical developed (TD) group.

In Experiment 7, previous responses biases found for TD (see Chapter 2) were replicated. A robust overshoot response bias was found when the last expression of the Joy and Anger videos was neutral and even when the last frame belonged to the same emotion category as the start emotion. In Experiment 8, the perceptual bias was completely removed by the change of the agent’s identity in the last neutral frame (i.e. the frame that had to be evaluated). This result ruled out the possible role played by contrast effects, as the contrast between the maximally happy or angry expression in the first frame and the neutral expression in the last frame was preserved. The change in identity did not affect this contrast, but it did affect emotional anticipation. The identity change manipulation also ruled out the RM explanation, thus replicating the findings reported in Chapter 2.

In Experiment 9, a mask (400 ms duration) was inserted just before the neutral expression in the last frame of the clip. The rationale was that the mask should reset those low-level processes that rely on the continuous flow of motion (cf. Enns et al., 2010). As the mask insertion did not affect the overshoot bias, contributions of extrapolations of curves from the face have to be rejected.

The looped video-sequence in Experiment 10 specifically tested for contrast and RM effects. TD individuals judged the last neutral expression of the Neutral-to-joy-to-neutral and Neutral-to-anger-to-neutral loops as slightly happy and slightly angry, respectively. This bias was opposite in direction to the normal overshoot effect, and was therefore called an ‘undershoot’ effect. The video-clips in the loop condition may have better resembled everyday social interactions (more ecologically-valid) than other video-sequences used in our studies. The observer might have processed the agent’s brief smile in the Neutral-to-joy-to-neutral videos as a positive social signal toward him/her, and the frown in the neutral-to-angry-to-neutral videos as a negative social signal. Consequently, a positive attribution may have been made to the agent in the Neutral-to-joy-to-neutral condition and a negative attribution in the Neutral-to-anger-to-neutral condition, resulting in the undershoot bias in the evaluations. In future research this interpretation could be tested further by manipulating the eye gaze direction of the agent. It is possible that the rapid change in the emotional facial expression in combination with averted eye gaze may not convey the positive or negative signal towards the observer (as the agent’s attention would be directed elsewhere) and would therefore not generate the undershoot response bias.

The finding of an ‘undershoot’ effect gives rise to two conclusions. First, contrast effects did not play a major role as we did find a response bias. If the overshoot of Experiment 7 was due to the contrast between the ‘full blown’ expression and the neutral one, then in the loop condition, where the two expressions in question were both neutral, we should not have found a bias. Second, if the overshoot in Experiment 7 was caused by RM, then we should have found an overshoot in Experiment 10 too (considering that the second half of the loop was identical to the video-clips shown in Experiment 7). In conclusion, on the basis of these results we can claim that the overshoot as found in TD individuals was not caused by purely bottom-up visual processes, but rather that it was generated by a top-down mechanism, involving an involuntary anticipation made by the observer about the agent’s emotional state of mind in the immediate future.

In summary: the results found with TD individuals in the tasks presented in Experiments 7, 8, 9 and 10 replicated previous findings of Experiments 1, 2, 3 and 5 reported in Chapter 2.

The Asperger’s syndrome (AS) group.

In Experiment 7, the AS group showed the same strong overshoot response bias as the control group. This might suggest they used the same mechanism, i.e. emotional anticipation, or, alternatively, it could mean they used a different strategy which led to the same response outcome. Experiments 8, 9 and 10 have shed light on the underlying mechanism which might have operated in the AS group.

In Experiment 8 the Identity-change manipulation did not remove the overshoot, while in the TD group it did. The presence of the bias in this condition cannot be accounted for by a difficulty in detecting the identity-change, as the extent of overshoot was quite similar between the trials where the identity-change involved a gender swift and the trials where the gender of the two identities was the same. The overshoot in the identity change condition suggests that contrast effects or extrapolation (RM) might have constituted the alternative route AS individuals used to evaluate the expressions in the ‘emotional anticipation’ task.

Experiment 9 ruled out the possibility that AS individuals have processed the stimuli by focusing on details and extrapolating curves in the face, as the overshoot response bias was still present after inserting the mask to interrupt the motion flow of the sequence.

Finally, the Loop condition in Experiment 10 clarified the roles played by contrast and RM effects. AS participants evaluated the last neutral expression as neutral, i.e. identical to the starting expression. This suggests that the alternative mechanism which played a major role in bringing about the overshoot bias in AS individuals is the contrast between the initial and the last frames (each presented for 300 ms) of the video-sequence (see Tanaka-Matsumi et al., 1995; Thayer, 1980). The lack of a perceptual bias in Experiment 10 suggests that, similarly as was found for the TD individuals, the hypothesis based on RM has to be rejected for the AS individuals. RM would have resulted in an overshoot effect, i.e. an extrapolation of the second part of the video sequence. RM effects were found for ASD individuals by using video-clips which were the reverse of ours (neutral-to-emotion; Uono et al., 2010). However, it should be noted that the neutral-to-emotion video-clips employed in the study by Uono et al., (2010) do not involve contrast effects, whereas these effects may occur in the overshoot paradigm, where the emotional expression could have influenced the perceptual evaluation of the target neutral face. Therefore, the findings reported in Chapter 3 confirm that the overshoot bias does not seem to be generated by RM effects.

In summary: AS participants showed an ‘overshoot’ bias similar to TD individuals in Experiment 7. However in subsequent manipulations AS participants did evaluate the video-clips differently from TD participants. Specifically, in the IC manipulation (Experiment 8) the overshoot did occur with AS group, leaving open the possibility that the perceptual distortion was due to RM or contrast effects. The Mask manipulation (Experiment 9) had no effect on the overshoot, thus minimizing the role played by the extrapolation of literal facial features from the flow of motion. Finally, in the Loop manipulation the absence of a significant ‘undershoot’ effect excludes RM either on a posite-negative valance dimension or on facial features. Therefore, the most likely mechanism causing the overshoot bias with AS participants seems to rely on contrast effects.

Which alternative mechanisms might be employed by individuals with AS to evaluate the expressions in the emotional anticipation task?

People with AS are more inclined to understand human actions in a systemizing manner, by representing them in terms of their physical characteristics rather than their mental characteristics (Baron Cohen, 2002). Overall our findings suggested that AS participants could have relied on the low-level visual characteristics of dynamic facial expressions, rather than on the associated emotional attributions. In contrast, TD individuals tend to involuntarily extract these emotional attributions from the perceptual history pertaining to that particular agent. This is in line with the evidence that AS individuals may anticipate other’s actions by using atypical strategies, showing a lack of comprehension of the underlying psychological meaning (Hudson et al., 2011). The current findings are also in accordance with the proposal that AS individuals may process the agentive social stimuli in the same way as they would process non-social/agentive objects (Baron Cohen, 2002; Jellema & Perrett, 2002a,b; 2007), which would explain their impairments in social and communicative competencies and empathy (Baron-Cohen, 2006, Driver et al., 1999; Ristic et al., 2005).

An alternative hypothesis that may go some way in explaining the deficits in social cognition and emotion reading in AS individuals is the Weak Central Coherence theory (WWC) theory (Happé, 1999; Happé and Frith, 2006). The WWC model claims that difficulties of ASD individuals in processing facial expressions of emotions are due to an enhanced attention to local details of the stimuli, thereby losing sight of the global configuration (Humphreys et al., 2007). Focusing on local features, such as the U-shaped mouth in the happy expression, might make them more prone to apply an extrapolation of these features. However, in the Mask condition (which should disrupt extrapolation on the basis of motion flow; cf. Enns et al., 2010) the AS individuals did show a normal overshoot response bias, suggesting that they did not applied an extrapolation.

Moreover, our findings are in line with some evidence that the visual processing of facial expressions seems to be preserved in HFA/AS participants (Ogai et al., 2003; Rutherford & Towns, 2008; Wicker et al., 2008), especially when dynamic stimuli presentations are employed (Loveland et al., 1997; Gepner et al., 2001). If there was a deficit in visual processing, or a lack of attention for this particular class of stimuli, then AS participants would not have produced the robust response bias in Experiment 7. We can also exclude the possibility that AS participants might not have been able to identify the gradually less intense expressions during the morph sequence. A recent study using dynamic facial expressions depicted at different intensities showed that adolescents with HFA found it difficult to accurately detect low and medium intensities (20 - 70%) of anger, but not of happy expressions, which were correctly recognized also at lower intensities (Law Smith et al., 2010). The overshoot bias in the current study was found for both the Joy-to-neutral and Anger-to-neutral videos. Overall the overshoot bias in AS individuals appeared somewhat less enhanced for the anger videos. However, this trend should be considered with caution as the calibrated scores of the neutral expressions cannot be treated as absolute measures. Further, we used individuals with AS, which were indeed very mildly affected (ADOS scores between 7 and 9), while the Smith et al. (2010) study used individuals with HFA, which are more severely affected.

It would be interesting to examine whether ASD individuals with more severe symptoms than those with AS might be even more prone to apply alternative rules, or adopt other compensatory strategies, to try and make sense of other people’s emotional states from social cues.

3.6.1 Possible neurophysiological dysfunctions supporting the hypothesis of the lack of emotional anticipation in ASD

We previously proposed that the neural mechanism underlying emotional anticipation in TD individuals could involve mirror mechanisms (Palumbo & Jellema, under review). The Mirror Neuron Mechanism (MNM), which is provided with a description of the observed facial dynamics by the superior temporal sulcus (STS), might underlie the tacit/intuitive understanding of others’ facial expressions through motor simulation. This might shape (top-down) the perception of the visual stimulus in TD individuals. Emotional anticipation can be seen as a low-level mind-reading mechanism (Goldman, 2006; Goldman & Sripada, 2005), which does not involve any cognitive inferential processes or deliberate reasoning.

Several functional neuroimaging studies on ASD individuals have reported abnormal activity in the social brain areas, which include the fusiform face gyrus (Critchley et al., 2000; Pierce, Muller, Ambrose, Allen, & Courchesne, 2001; Wang, Dapretto, Hariri, Sigman & Bookheimer, 2004), the amygdala (Adolphs et al., 2001; Baron-Cohen et al., 2000; Ogai et al., 2003) and the posterior STS (Castelli et al., 2002; Pelphrey et al., 2007). This network, linking the structural processing of facial features with emotional expressions and biological motion, is supposed to be crucial in attributing emotional meaning to social events. In addition, it has been hypothesized that social impairments in ASD may be related to alterations in the Mirror Neuron System (MNS) (consisting basically of premotor cortex, inferior frontal gyrus and the posterior parietal cortex; see Rizzolatti & Craighero, 2005; Iacoboni et al., 1999), which would be responsible for a lack of imitation, ToM and empathy (Baron-Cohen, Avenanti, Walsh, & Aglioti, 2009; Dapretto et al., 2006; Iacoboni & Dapretto, 2006; McIntosh, Reichmann-Decker, Winkielamn, & Wilbarger, 2006; Minio-Paluello, Nishitani, Anikainen, & Hari, 2004; Rizzolatti & Fabbri-Destro, 2009; Williams, Whiten, & Singh, 2004; Williams et al., 2001).

Abnormally correlated activity between visual regions and STS or inferior frontal gyrus, and between fusiform face gyrus and amygdala, has been reported in ASD while observing emotional facial expressions or bodily actions (Castelli et al., 2002; Grèzes, Wicker, Berthoz, & de Gelder, 2009). Further, ASD individuals may lack internal representations of body states associated with actions and emotions (i.e. a lack of embodied simulation; Gallese, 2007), and this would preclude a direct experiential understanding of others, or 'intentional attunement' (Gallese, 2006a, 2006b).

Our current findings with AS individuals suggested that they possibly used alternative mechanisms to compensate for the absence of emotional anticipation. This seems to be in line with the 'Broken Mirror' hypothesis (Williams et al., 2001). This model assumes that the MNM uses 'direct matching' (Rizzolatti et al., 2001), which is that the observation of an action directly activates mirror neurons that (subthreshold) duplicate the motor program which would be needed to perform that action. In other words those neurons enable an unmediated transformation of the visual information into a motor code. This transformation can be seen as an offline embodied simulation for understanding the goal of that action. However, the role played by '*direct matching*' mirroring in the social impairments observed in ASD individuals is still debated (Hamilton, 2009; Hamilton, Brindley, & Frith, 2007; Marsh & Hamilton, 2011; Press, Richardson, & Bird, 2010; Sebanz, Knoblich, Stumpf, & Prinz, 2005; Southgate, Gergely, & Csibra, 2010; Southgate, & Hamilton, 2008; Spengler, Bird, & Brass, 2010).

One challenging alternative to the 'direct matching' mirroring hypothesis is the '*action reconstruction*' account proposed by Csibra (2007). This author emphasizes that action mirroring has a predictive function, which is generated by an interpretation of the action's goal. Therefore, according to Csibra's model, we first understand the meaning of an observed action and this meaning (input of mirror mechanism) would be 'simulated' in our motor system in order to generate a prediction.

In line with Csibra's model, the 'mentalizing' theory suggests that a specific cognitive deficit in 'high level' reasoning and in representing mental states can better explain the core of the social deficit in ASD (Frith, 2003; Frith et al., 1991). In support of this position an abnormal activation of the ‘mentalizing’ neural circuit, including medial prefrontal cortex and temporo-parietal junction (Brass et al., 2007) was found in ASD. However, it should be noted that the ‘constructive matching’ approach, as well as the hypothesis of a dysfunction of the mentalizing network, are in particular based on experimental paradigms employing manipulations of goal-directed movements executed with limbs, which do not directly tap into the social-emotional dimension.

3.6.2 Conclusions

The current study suggests that AS participants could have relied on visual perceptual characteristic of the presented social stimuli, rather than on the conveyed emotional significance, which was automatically grasped by TD participants. Thus, it is proposed that participants with AS solved the emotional anticipation task by using an atypical mechanistic, rule-based strategy, typically applied to understand non-agentive events, as a substitute for their impairment in emotion-reading. The results presented in Chapter 3 are in line with the evidence of compensatory mechanisms for action anticipation in AS individuals (Hudson et al., 2011). This may explain anecdotal ‘misunderstandings’ when interacting with others (cf. Frith, 1995), but also shows that people with mild autistic symptoms can compensate for the lack of social understanding by using refined cognitive skills. Further investigations are needed to clarify the neural mechanisms underlying emotional anticipation and its dysfunction.

CHAPTER 4

Emotional anticipation with ‘fake’ facial expressions

4.1 Introduction

In Chapter 2 individuals with typical development (TD) evaluated the neutral expression at the end of Joy-to-neutral and Anger-to-neutral video-clips as slightly angry and slightly happy respectively and showed an overshoot response bias. Further investigations were conducted to try to clarify the nature of the overshoot phenomenon. A series of six experiments presented in Chapter 2 ruled out the hypothesis that the overshoot effect was generated by basic visual processes such as contrast effects, adaptation and extrapolation or Representational Momentum (RM). Alternatively, it was suggested a different mechanism, called emotional anticipation, would better account for the perceptual distortions. That is, participants with TD were able to involuntary anticipate the emotional state of the actors based on the perceptual history presented immediately prior to the target neutral expression. Thus, the anticipated negative emotional state of the actors at the end of the Joy-to-neutral videos and the anticipated positive state at the end of the Anger-to-neutral videos in turn modulated (top-down) the perceptual evaluation of the last neutral frame (Palumbo & Jellema, under review).

Next, in a series of four experiments presented in Chapter 3, the emotional anticipation hypothesis was tested on individuals with Asperger syndrome (AS), who may have difficulties in emotion-reading and may have less empathic reactions to emotional stimuli (Baron-Cohen, 2006). AS individuals did show the same perceptual distortion as TD individuals (see Experiment 7, Chapter 3). However, it was found that this response bias in AS individuals was not generated by emotional anticipation rather by possible mechanistic compensatory strategies (see Experiment 10, Chapter 3).

Taken together these findings may suggest the existence of an automatic low-level emotion reading mechanism (Goldman, 2006; Goldman & Sripada, 2005), which enables people with TD to generate immediate expectations about how an emotional state of another individual may develop in the immediate future, thus providing a direct tool to react promptly and appropriately in social interactions. In contrast, individuals with AS, who lack this direct and automatic access to the other person's mental state, may possibly develop alternative mechanisms to face the social world, even though with less immediacy and accuracy.

4.1.1 Not all smiles have the same meaning

Facial expressions are powerful signifiers of others' affective states (Darwin, 1872/1998; Ekman, 2004), but they are also complex social stimuli. Facial expressivity does not always deliver a clear emotional message as it also serves different social, communicatory functions. A Duchenne-smile is universally recognized as expressing happiness (Ekman, 1971; Ekman & Friesen, 1971, Izard, 1994). However smiles, even when they are expressed spontaneously, may assume different meanings according to a variety of social norms and contexts as it is the case of 'greeting' smiles (Eibl-Eibesfeld, 1972) or 'sardonic' smiles reported by Darwin (1872, p. 251). A smile may also express amusement, interest, compassion, respect or embarrassment (Keltner 1995; Hess, Beaupre, & Cheung, 2002). Ekman (2001) presented a total of 18 main smile types, and 50 subtypes, each one with a different social value, but not necessarily related to an enjoyment experience (see also Cashdan, 2004; Fridlund 1991, 2002).

One of the main distinctions between smiles is whether they are genuinely (spontaneously) expressed or whether they are deliberatively (voluntary) posed (Ekman, 2003). There is evidence that people can detect differences in smile type (Frank, 2002; Frank, Ekman, & Friesen, 1993). This ability starts developing early in childhood (Gross and Harris, 1988; Gosselin, Perron, Legault, & Campanella, 2002; Pons, Harris, & de Rosnay, 2004; Del Giudice & Colle, 2007) and is important to cope with the complexity of the social world.

4.1.1.1 Genuine versus 'false' smiles

Enjoyment smiles are spontaneous signals of felt emotion while posed non-enjoyment smiles may hide a neutral state of mind (phony smile; Ekman & Friesen, 1982) or mask a negative feeling such as disgust (Ekman, Friesen, & Ancoli, 1980), disappointment (Kraut & Johnston 1979) or sadness and uncertainty (Klineberg, 1940). Almost all studies emphasized perceptual differences related to the morphology and dynamics of the genuine and posed smiles (Ekman, Hager, & Friesen, 1981; Frank & Ekman, 1993). In the research carried out by Ekman and collaborators with the Facial Action Coding System (FACS; Ekman & Friesen, 1978), different facial areas (or units) involved in the felt and unfelt smiles were identified (see also Gosselin, Kirouac, & Dorè, 1997; Kohler et al., 2004).

A genuine smile presents the Duchenne marker (Action Unit 6, AU6), namely the contraction of the external strand of the muscle Orbicularis Oculi around the eye, which generates a lifting of the cheeks and the “crow’s feet” wrinkles at the corners of the eyes (Ekman & Friesen, 1982). Tense lower eyelids, raised cheeks and lip corners pulled up by the Zygomaticus Major contraction can also be indicative signs of positive emotion (Frank et al., 1993; Soussignan 2002). In contrast, a posed smile can be detected on the basis of lip corners asymmetry (Ekman & Friesen, 1982; Ekman et al., 1981) and the lack of contraction of the Orbicularis Oculi, which prevents the cheeks to rise up to the same extent as they would in case of a genuine smile.

A non-Duchenne smile is visibly evident by the absence of the wrinkles around the eyes. Importantly, studies with EMG reported that different muscles of the face are involved in genuine and posed smiles. For instance the Zygomaticus Major is more active during genuine smiles than during fake ones (Tassinary & Cacioppo, 1989). Moreover it is suggested that the Risorious and Buccinator muscles would be more involved while faking a smile (Oberman, Winkielman, & Ramachandran, 2007). These latter are the “say cheese” muscles. An illustration of the different muscles of the face involved in a smile is presented in Fig. 4.1).

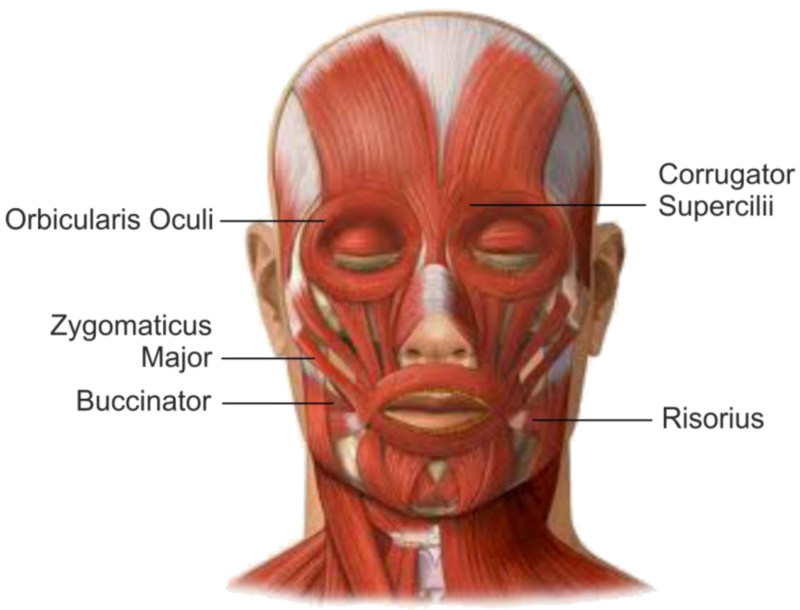

Figure 4.1. Illustration of the facial muscles. The Zygomaticus Major and Orbicularis Oculi (left side) are typically activated during genuine smiles and the Corrugator

Supercilii (right side) is responsible of frowning during angry expressions. The Buccinator (left side) and Risorius (right side) are thought to play a major role in the production of fake smiles.

In parallel to studies which mainly focused on the difference between felt and unfelt smiles on the basis of Duchenne markers, other investigations highlighted the importance of timing and dynamicity (see Abel et al. 2002, for a review). Deliberate, posed smiles feature more irregular actions, characterized by more phases, pauses, and stepwise intensity changes than spontaneous smiles which are smoother, and also have slower onset and offset durations (Ekman & Friesen, 1982; Hess & Kleck, 1990; Krumhuber, Manstead, & Kappas, 2007; Schmidt, Ambadar, Cohn, & Reed, 2006).

People showing enjoyment smiles are judged as happier (Otta, Abrosio, & Hoshino, 1996; Otta, Lira, Delevati, Cesar, & Pires1994), more attractive, polite and truthful (Hess et al., 2002; Reis et al., 1990) and are liked more (Young & Beier, 1977; Palmer & Simmons, 1995). Posed smiles are highly adaptive because they are assumed to deliberatily communicate social intentions and regulate interpersonal relationships. However, being ‘non felt’ and artificial, as opposed to spontaneous and natural, they may appear as less pleasant, thus producing less emphatic reactions (Surakka & Hietanen, 1998).

4.1.1.2 Can negative facial expressions be voluntarily posed?

Almost all of the literature on deceptive facial expressions is focused on felt and unfelt smiles, while few investigations were carried out on the deliberate use of other facial expressions. It is reported that negative emotions (sadness and fear) were more difficult to falsify than happiness (Porter & ten Brinke, 2008). Another study focused on detecting deceptive pain from facial displays (Hill & Craig, 2002). There is substantial agreement on the difficulty to identify the deceptive cues associated with pretended negative expressions (Hurley & Frank, 2011). Moreover, little is known about the temporal properties and contiguity of specific facial actions involved in the production of such expressions.

It would be interesting to extend research into this area to better understand the social function of posed negative facial expressions and the underlying mechanism which enable the understanding of the social meaning conveyed by those expressions.

4.1.2 Embodiment of facial expressions

Relying on theories of embodied cognition in the recognition of facial expressions (Carr et al., 2003; Goldman, 2006; Jabbi et al., 2007; Niedenthal, 2007; Pitcher et al., 2008, Wicker et al., 2003), Niedenthal, Mermillod, Maringer & Hess (2010) proposed the simulation of smiles model (SIMS). This model ascribes an important role to embodied simulation in understanding the signal (social or emotional) linked to the facial expressivity (Gallese, 2007; Goldman & Sripada, 2005). According to the SIMS model the visual input, arising from specific regions of the face associated with the production of a particular emotional expression, elicits a congruent emotional response in the observer. Thus, the internal simulation of the observed facial expressions is promoted by the automatic facial motor mimicry (Niedenthal, 2007). It is well known that facial expressions of negative and positive affect produce rapid, unconscious and specific facial EMG responses over Corrugator Supercilii and Zygomatic Major muscles, respectively (Dimberg & Thunberg, 1998;

Dimberg et al., 2000; Larsen et al., 2003). Such mimicry reactions may reflect activity in the neural mechanisms that play a role in the tacit understanding of the emotion conveyed by the face (Atkinson & Adolphs, 2005; Pitcher, Garrido, Walsh, & Duchaine, 2008). Therefore, individuals can distinguish happy from angry expressions on the basis of different somatosensory and affective experiences that those expressions produce in them. Interestingly, in a recent work by Maringer, Krumhuber, Fischer, & Niedenthal (2011) it is suggested that those somatosensory simulations may also help the perceiver to discriminate between genuine and false facial expressions. That is, a smile will be judged as a genuine expression of happiness if it induces such a feeling in the observer. The simulation of a false smile may be reflected by weaker mimicry (Surakka & Hietanen, 1998), or by a mimicry which may involve different facial muscles (Oberman et al., 2007), as compared to genuine smiles. The result is that the smile is experienced as less positive by the perceiver, and therefore judged as unfelt or false. At the present the SIMS model is applied only to smiling faces, while no attempts were made with different facial expressions, in particular with displays of negative facial affect.

4.1.3 Aims of this study

Facial expressions can have an emotional meaning but they can also be unaffective social signals (Fridlund, 1994). Successful interaction and communication implies the ability to use and interpret these subtle differences correctly. For instance, interpersonal communication sometimes can be voluntarily aimed to manipulate or convince the interlocutor, which is not uncommon in domains such as forensic, marketing, business, entertainment and politics. These are all examples suggesting that anticipating other people's behavior may help to face the complexity of the social world successfully, especially when deceptive signals are employed.

In the previous studies (Chapters 2 and 3) it was shown that rapid changes in the intensity of emotional facial expressions (Joy-to-neutral and Anger-to-neutral) generated an automatic anticipation of the final emotional state of the agent in TD observers. Thus, the perception of the neutral expression depicted at the end of the happy and angry sequences was biased in the direction of the anticipated state of mind. In Chapters 2 and 3, emotional anticipation was proposed to be a low level mindreading mechanism underpinning embodied simulation of the agent's dynamic facial expression (Gallese, 2007; Goldman, 2006; Goldman & Sripada, 2005), in line with SIMS model (Niedenthal et al., 2010). The main question addressed in Chapter 4 is how participants would perform on the 'emotional anticipation' task if the video-clips would display unfelt, deliberately faked facial expressions. Such faked expressions, contrary to genuine expressions, do not have a clear relationship to the emotional state of the actor. Hereto, a new set of stimuli was created, in which each actor displayed both felt, genuine and unfelt, posed (or commonly termed 'fake') happy and angry facial expressions. Despite genuine and fake expressions being relatively similar in terms of facial features, the conveyed meaning reflects different mental states (emotion vs. social intention). Therefore, the 'emotional anticipation' hypothesis would predict a significant reduction of the overshoot bias with fake facial expressions as compared to genuine ones.

Moreover, Maringer and colleagues (2011) showed that only when blocking the observer's mimicry, thus inhibiting embodied simulation of the facial expression, the context and stereotypes or beliefs are taken into account in the evaluation of the smile's meaning. In contrast, when facial mimicry was active, no difference was reported in judgments of smile genuineness as a function of the context in which the smile was embedded. The authors concluded that the observer's interpretation of another's smile on the basis of a re-enactment of the other's affective state was strong enough to override context information.

The additional question addressed in Chapter 4 is whether, and how, participants' explicit knowledge about the lack of authenticity in the emotional perceptual history (Joy-to-neutral and Anger-to-neutral) would modulate the perceptual evaluation of the last neutral expression. Does explicit knowledge about the genuineness of facial expressions influence anticipatory mechanisms?

4.3 Experiment 11

A first experiment was conducted to test whether the overshoot would be replicated by using the new set of genuine happy and angry facial expressions. Short video-clips were created as reported in Chapters 2 and 3. According to the 'emotional anticipation' hypothesis the prediction was to find an overshoot effect with the new stimuli as the faces were all depicting genuine emotional expressions. The overshoot bias with Ekman's set of pictures (Ekman & Friesen, 1976) was replicated consistently across all the experiments in Chapters 2 and 3. Therefore, the new set of stimuli should induce a similar overshoot response bias if there were to be used in subsequent experiments.

4.3.1 Method

4.3.1.1 Participants

Nineteen undergraduate Psychology students at Hull University (UK) took part in the experiment. After applying exclusion criteria (see data reduction below), the data of fifteen participants were included in the analysis (age, M = 19.7 years, SD = 1.7 years; 14 females; 1 male). None of the participants took part in any of the other studies. All participants had normal or corrected-to-normal vision, and provided written consent prior to the experiment. Participants received course credits or a fee for taking part.

The study was approved by the Ethics committee of Hull University.

4.3.1.2 Stimuli

Pictures of genuine and posed facial expressions were realized in collaboration with four professional actors (2 females: CL, MA and 2 males: BR, LO) and one professional photographer of “Accademia Arvamus”, a cultural association in Rome, Italy.

Actors applied the Stanislavsky technique to try to experience the target emotion (joy and anger; Gosselin et al., 1995). Posed, deliberate happy and angry expressions were displayed by asking the actors to prevent facial muscles’ activity typically involved in enjoyment-smiles (Orbicularis Oculi and Zygomaticus Major) and in spontaneous angry expressions (Corrugator Supercilii). Next the method described by Duclos & Laird (2001) was used. Specifically, for the posed smiles actors were trained to relax the muscle around the eyes, as if they were posing a neutral expression, thus avoiding the natural contraction of the Orbicularis Oculi which typically occurs while smiling genuinely. Actors were instructed to maintain this position for 15 seconds while lifting the lip corners. For the posed angry expressions actors were trained to display an angry expression (i.e. by clenching the teeth or compressing the superior lips against the inferior ones) and maintain this position for 15 seconds without contracting the brow. Actors were also asked to display neutral expressions. Standardized editing procedures using Adobe Photoshop 8.0 were applied to selected pictures. Specifically, although the camera's settings and the ambient lights were fixed, luminance and contrast were equalized and the hair and the background were blackened. Finally, pictures were transformed in grey-scale bitmaps to avoid any possible visual effect related to skin color. The eyes of all actors were positioned on approximately the same location on the screen. Faces measured about 13 x 18 cm when displayed on the screen. The final pictures used in the morphing procedures are illustrated in Figure 4.2.

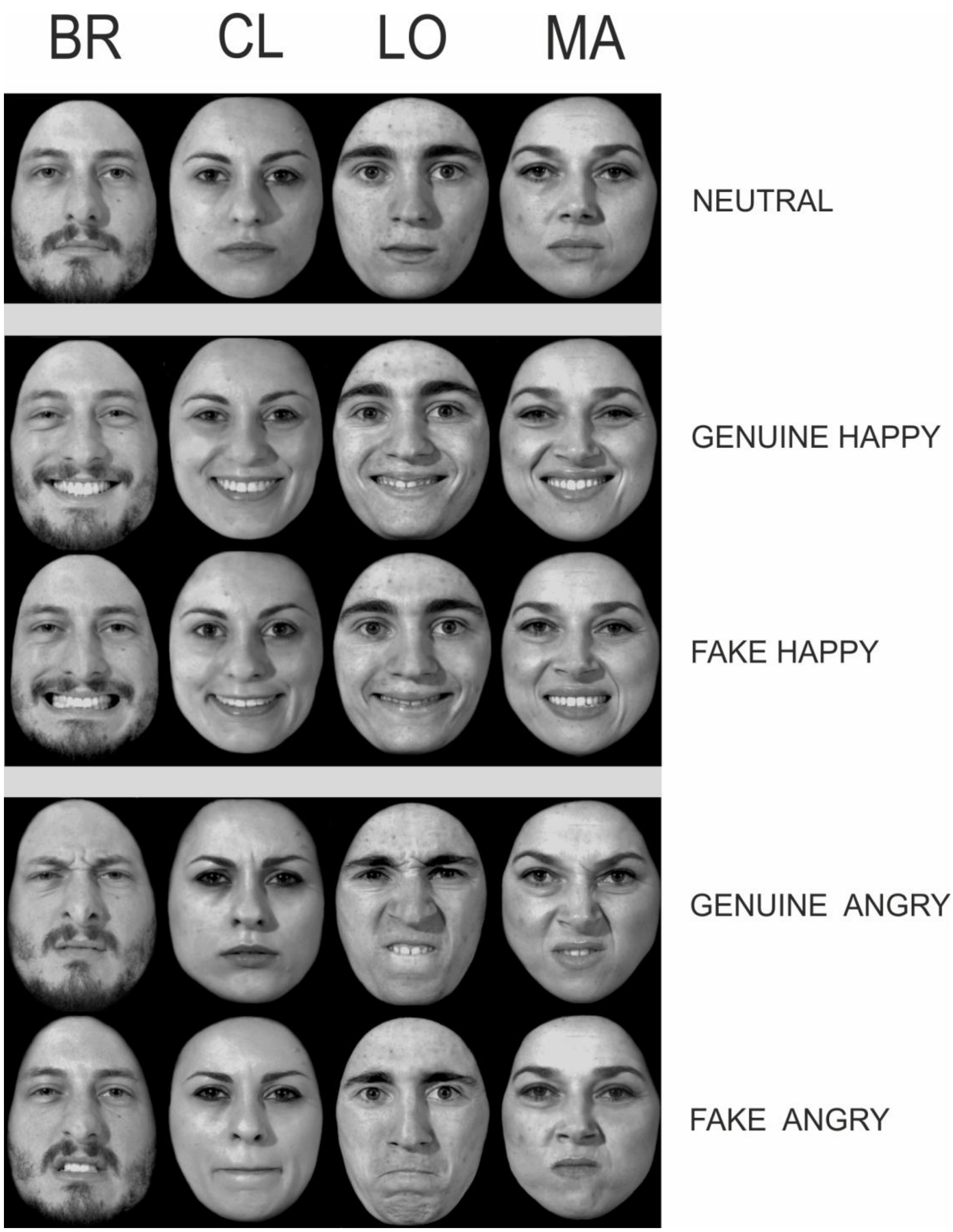

Figure 4.2. Illustration of the new set of stimuli with neutral, genuine and fake happy and angry facial expressions. At the top the initials of the four actors are indicated.

Similar to Chapters 2 and 3, video-clips, which consisted of nine interpolated images in between the full-blown genuine expression of joy or anger (called 100%) and the neutral expression (0%), were created at equal steps of 10% intensity change using multi-morphing software (Sqirlz morph 2.1, http://www.xiberpix.net/SqirlzMorph.html).

The first frame of each video sequence showed the emotional expression at 100% intensity and was presented for 300 ms, the subsequent interpolated frames were shown for 30 ms each (Figure 4.3). The total duration of the morph sequence was 270 ms (9 x 30 ms). The total duration of the stimulus presentation was 870 ms. These parameters were identical to those of the clips presented in Chapters 2 and 3.

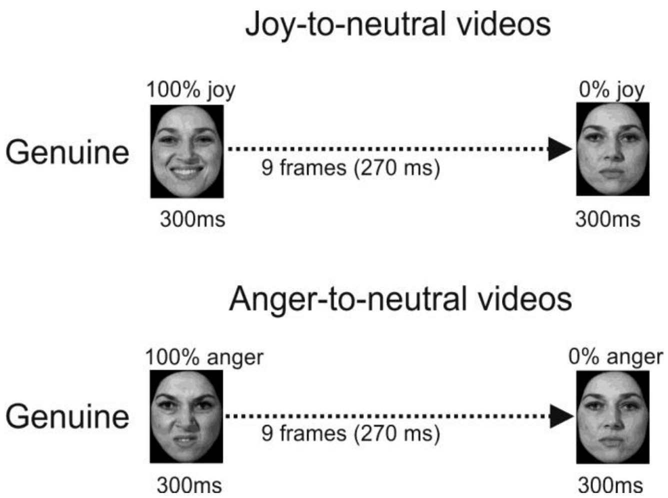

Figure 4.3. Schematic illustration of the stimulus presentations in Experiment 11. Joy-to-neutral videos started with a facial expression of intense joy, which gradually morphed into a neutral expression (top panel). Anger-to-neutral videos started with a facial expression of intense anger, which gradually morphed into a neutral expression (bottom panel).

4.3.1.3 Experimental procedure

The experimental procedure was the same as in the previous studies in Chapters 2 and 3. Participants were seated at a viewing distance of 80 cm from a PC screen (17-inch monitor, 1024 x 268 pixels, 100 Hz). The stimuli were presented using E-Prime (v. 1.2; Psychology Software Tools, Inc.).

First, participants calibrated the static neutral expressions of the new actors. Each calibration trial began with a black fixation cross on a grey background displayed in the center of the screen for 500 ms followed by the static face that was presented for 600 ms against a black background (each expression was presented twice, in random order, resulting in 8 trials). Participants were prompted to rate the calibration trials using the same 5-point scale as in Chapters 2 and 3 ranging from slightly angry (1) via neutral (3) to slightly happy (5), by pressing one of the 5 labeled keys on the SR-Box (Psychology Software Tools, Inc., USA).

After the calibration, 8 practice trials were completed (4 actors X 2 perceptual histories, joy and anger), followed by 48 randomized experimental trials (4 actors X 2 perceptual histories X 6 repetitions). Each trial consisted of a fixation cross displayed for 500 ms followed by the video-clip (870 ms). As in the calibration trials, participants were prompted to rate the last neutral expression of the sequence by using the same 5-point scale. Participants were instructed to respond within 3 seconds. The duration of the entire experiment was 20 minutes.

.

4.3.1.4 Data reduction and analysis

The same exclusion criteria as reported in the previous studies were applied. A 6.4% of the total number of trials were removed because RTs fell below 250 ms or above 3000 ms. Participants were excluded if more than 25% of the RT values were outside the 250 – 3000 ms range (n = 4). The ±2.5 SD rule to the mean difference ratings did not exclude participants.

The calibration scores were used to adjust the scores in the experimental trials; a calibration factor [3.00 minus the calibration score] was added to the experimental scores. This allowed performing one-sample t tests with test value 3.00 on the scores in the two perceptual histories. The calibration scores for the neutral expression of each actor are plotted in Figure 4.4. All statistical analyses were performed on the calibrated scores.

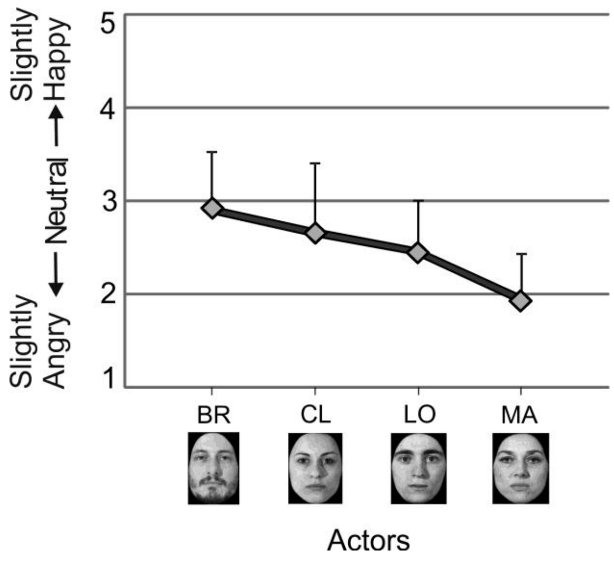

Figure 4.4. Calibration scores in Experiment 11. Ratings on the 5-point scale (vertical axis) for the static neutral expression. Scores are shown for each of the 4 actors (their initials are indicated on the horizontal axis). Error bars indicate +1SD.

4.3.2 Results

The results showed the usual overshoot effect (Figure 4.5). The neutral expressions at the end of the Joy-to-neutral sequences (M = 2.75, SD = 0.40) were evaluated as significantly more angry compared to the same neutral expressions at the end of the Anger-to-neutral sequences (M = 3.42, SD = 0.51) $t(14) = -6.31$, $p < .0001$). Further, the mean ratings for the Joy-to-neutral and the Anger-to-neutral videos differed significantly from 3.00 (joy-video: $t(14) = -2.45$, $p = .028$; anger-video: $t(14) = 3.22$, $p = .006$; one sample t-test, two-tailed).

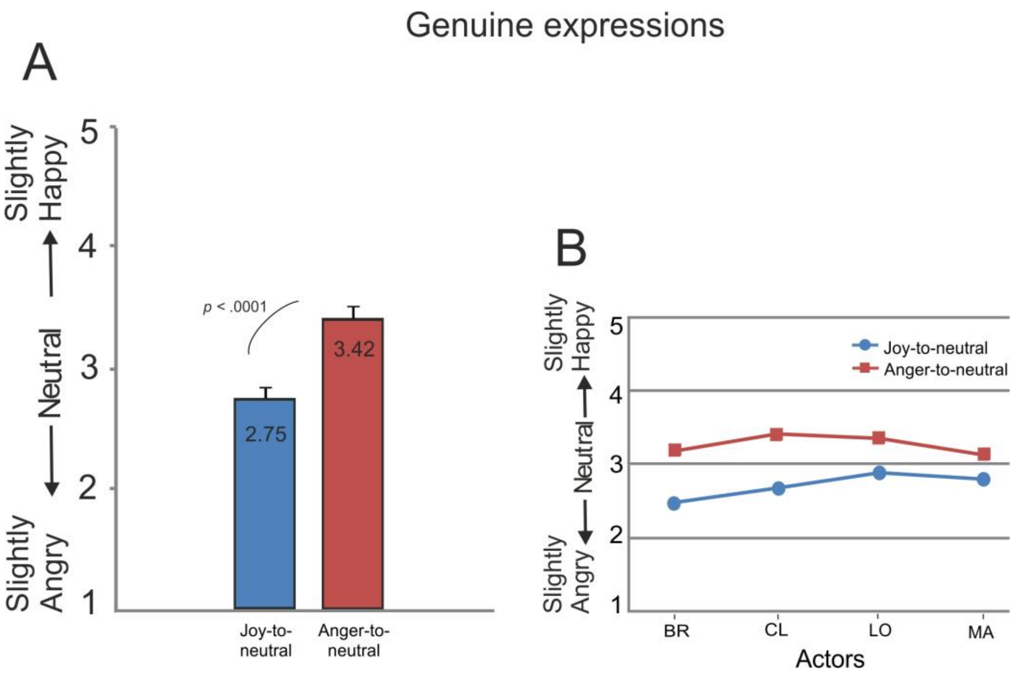

Figure 4.5. Results of Experiment 11. (A) Scores for the neutral expressions at the end of the genuine Joy-to-neutral and Anger-to-neutral sequences. Error bars indicate SEM. (B) Ratings for each of the four actors with genuine Joy-to-neutral and Anger-to-neutral videos are shown to illustrate the consistency across actors.

4.3.3 Discussion

The aim of Experiment 11 was to test whether the overshoot response bias could be replicated when using a new set of (genuine) facial expression pictures. A comparison between the results found with the actors of the Ekman and Friesen (1976) set (Experiment 1, Chapter 2) and those obtained with the new actors (Experiment 11, Chapter 4) revealed that the extent of overshoot bias with the new actors was very similar to that found with the Ekman & Friesen actors: $t(46) = .088$, $p = .930$ for Joy-to-neutral videos and $t(46) = .622$, $p = .537$ for Anger-to-neutral videos; independent sample t-tests. As the new set of facial expressions were comparable to the Ekman set in terms of induced overshoot bias, it was decided to use them in the subsequent experiments.

4.4 Experiment 12

In Experiment 12 the emotional anticipation hypothesis was tested further by directly comparing the genuine Joy-to-neutral videos with the fake Joy-to-neutral videos.

The rationale was that in case of the deliberate, unfelt expressions of joy (fake), the observer would not associate the change in the expression with a congruent change in the affective state of the actor. At the start of the clip (100% faked smile) the actor’s emotional state of mind was not happy (but maybe neutral or even negative), and therefore, in terms of emotional anticipation, one does not expect to find an overshoot response bias following the neutral expression at the end of the clip.

Therefore, if the overshoot bias, as reported in Chapters 2 and 3 with TD individuals, was due to an anticipation of the emotional state of the agent, then in the current experiment the prediction was to find a weaker overshoot response bias with the fake Joy-to-neutral videos as compared to the genuine ones.

Importantly, in Experiment 2 participants were naïve about the presence of any faked expressions! No mention of the possibility that expressions could be faked was made during the instructions prior to the experiment. This opened up the possibility that observers would simply not detect the fakeness of the smiles and would, by default, assume that all presented smiles were genuine. However, the aim was to explore whether the observers would be able to pick up on the fakeness by themselves. Thus, if the overshoot in the fake condition would not differ from that in the genuine condition, then that could be either due to participants not realizing that some smiles were faked, or to the fakeness not having an effect.

Video-sequences of fake angry expressions were not included. This was because deliberate, posed angry expressions are unusual in day-to-day life and possibly harder to detect without additional contextual information. For example, people can playfully pose an angry face while interacting with a child. Polite or situational smiles are routinely posed in everyday social interactions and can be recognized with less effort even without additional information.

4.4.1 Method

4.4.1.1 Participants

Twenty-seven undergraduate Psychology students at Hull University (UK) took part in the experiment. All participants were recruited applying the same restrictions as in Experiment 11.

After applying exclusion criteria (see data reduction below), the data of twenty-three participants were included in the analysis (age, M = 19.6 years, SD = 1.9 years; 18 females, 5 males). None of the participants took part in any of the other experiments.

4.4.1.2 Stimuli

Pictures depicting genuine and fake happy facial expressions of the same four actors were used (see Figure 4.2).

A preliminary test was conducted with 45 participants (30 females, 15 males; age: M = 21.3, SD = 2.2) to assess whether they could distinguish between the static genuine and fake smiles presented on a black background for 500 ms. The total number of trials was 48 (4 actors X 2 expressions X 6 repetitions), randomized in one block. The task was to press key ‘Z’ for genuine and key ‘M’ for fake on a keyboard (the two keys were labeled as ‘G’ and ‘F’ respectively to avoid confusion). Participants detected genuine and fake smiles above chance. The mean of correct responses with genuine expressions was .72, SD = .13 which was significantly different from .50 ($t(44) = 11.70$, $p < .0001$, one sample t-test). A similar result was obtained for the mean of correct responses with fake expressions: .75, SD = .17 ($t(44) = 9.71$, $p < .0001$).

Creation of the video-sequences was as in Experiment 11. The Anger-to-neutral videos where not included at this stage as the fake angry expressions are less known and they may be more difficult to detect without additional contextual information. An example of the stimulus presentation is illustrated in Figure 4.6. In order to enable participants to better detect the type of smile (genuine or fake) depicted in the first frame of the video-clip, the duration of the first frame was set to 600 ms. This choice was constrained by the ‘overshoot’ paradigm. It was previously reported that a change of the first frame duration up to 600 ms would not modulate the bias, though it cannot be excluded that longer durations might give rise to adaptation and aftereffects (Experiment 3, Chapter 2). Moreover, the minimum amount of time required to ensure the detection of false smiles is unknown; most studies used real-time video recordings of the whole expression (onset-apex-offset; see Ekman & Friesen, 1982), or very long exposures (8 s; Surakka & Hietanen, 1998). The entire duration of the video-clips was 1170 ms (first frame 600 ms, last frame 300 ms and 9 interpolated frames 270 ms).

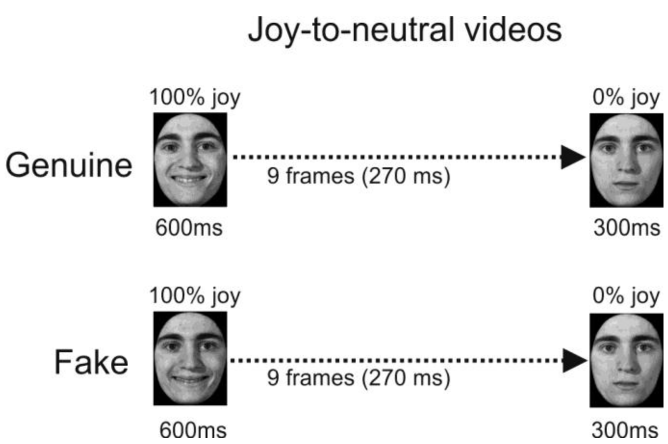

Figure 4.6. Schematic illustration of the stimulus presentations in Experiment 12. Joy-to-neutral videos started with a facial expression of intense genuine joy (600 ms), which gradually morphed into a neutral expression.

4.4.1.3 Experimental procedure

First, participants calibrated the static neutral expression of each actor, which were presented for 600 ms (8 trials: 4 actors X 2 repetitions). After the calibration phase, participants started the experimental session.

First, 8 practice trials were completed (4 actors X 2 expressions, genuine and fake), followed by 48 randomized experimental trials (4 actors X 2 expressions X 6 repetitions). Each trial consisted of a fixation cross displayed for 500 ms followed by the video-clips (1170 ms). Participants were prompted to rate the last neutral expression of the sequence using the same scale. Participants were instructed to respond within 3 seconds. The duration of the entire experiment was 20 minutes. All other aspects of the procedure were identical to Experiment 11.

4.4.1.4 Data reduction and analysis

Trials in which the RTs fell below 250 ms or above 3000 ms were removed (5.2% of total number of trials). Participants were excluded if more than 25% of the RT values were outside the above range (n = 2). The $\pm$2.5 SD rule to the mean difference ratings excluded two participants. The calibration scores for the neutral expression of each actor are shown in Figure 4.7. All statistical analyses were performed on the calibrated scores.

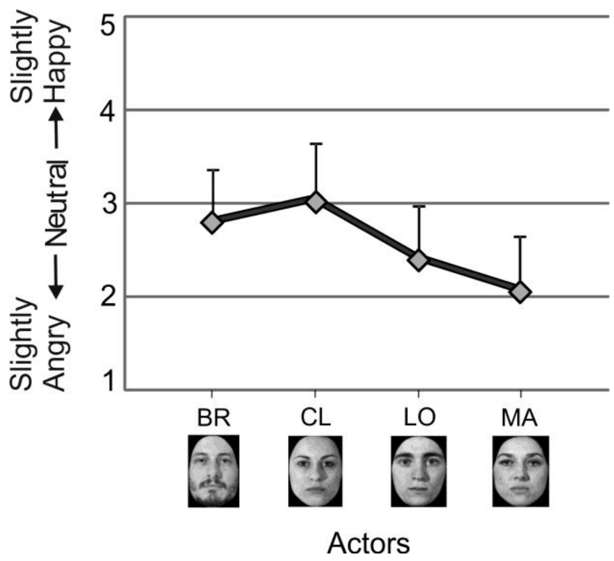

Figure 4.7. Calibration scores in Experiment 12. Ratings on the 5-point scale (vertical axis) for the static neutral expression. Scores are shown for each of the 4 actors (their initials are indicated on the horizontal axis). Error bars indicate +1SD.

4.4.2 Results

The results are illustrated in Figure 4.8. The mean ratings for the neutral expressions at the end of the genuine sequences (M = 2.80, SD = 0.49) and at the end of the fake ones (M = 2.85, SD = 0.36) did not differ from each other, paired sample t-test (t(22) = -6.94, p = .495).

One sample t-tests revealed that the ratings of the neutral faces at the end of the Joy-to-neutral videos in both the genuine and fake conditions just did not differ significantly from 3.00 (genuine: t(22) = -1.94, p = .065; fake: t(22) = -2.03, p = .055).

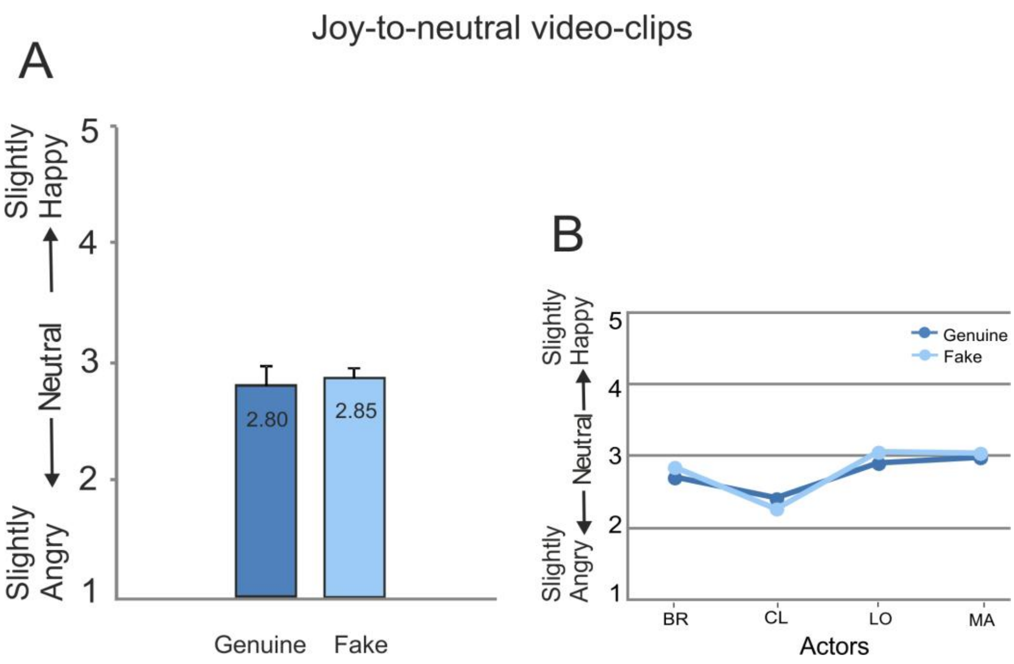

Figure 4.8. Results of Experiment 12. (A) Scores for the neutral expressions at the end of the Joy-to-neutral (horizontal axis) in the genuine and fake conditions. Error bars indicate SEM. (B) Ratings for each of the four actors with genuine and fake Joy-to-neutral videos are shown to illustrate the consistency across actors.

4.4.3 Discussion

In Experiment 12 video-clips of unfelt smiling expressions, which showed a gradual decrease in the emotion intensity, were used to test their effect on the overshoot response bias. In this experiment participants did not know that half of video-clips depicted fake smiles and the other half were genuine. Nevertheless, it was interesting to test whether they would pick up on the fakeness by themselves. As the results showed no difference between the genuine and fake conditions, there is no way to tell from this experiment whether they did or did not pick up on the fakeness.

It could be that the participants did not realize some expressions were faked, or it could be that they did realize it but that the fakeness had no effect. However, debriefing of participants suggested that they had no idea that some expressions were faked.

It should be noted that the absence of the Anger-to-neutral condition may have affected the results, in that it made of an overshoot bias for more difficult to verify. Normally the difference between the Joy-to-neutral and Anger-to-neutral conditions is the crucial measure, but in Experiment 2 it was only possible to rely on the difference with respect to the value '3.00' in one-sample t-tests. This inevitably introduces some error.

In Experiment 12 the first frame (full-blow expression) of the video-clip was presented for 600 ms, however, in Experiment 12 participants may have found more difficult to realize implicitly the difference between genuine and fake smiles embedded in the dynamic offset (the gradual extinction of the expression). Participants may have realized that the two video-clips were displaying different kinds of smiles, but this does not mean they understood these smiles were specifically genuine or fake (this was confirmed by debriefing participants).

It should be considered though that faked smiles do not necessarily need to be negative. People pose smiles in their every-day life. Especially polite or situational smiles are displayed automatically during interpersonal interactions (Eibl-Eibesfeld, 1972). Thus, even if these smiles are not linked to joy, they still may produce a positive reaction in the observer. Overall the fake smile has a strong adaptive function that is to facilitate interpersonal interactions. The fake smile is also considered a powerful tool to impress people positively.

4.5 Experiment 13

In Experiment 13, the labels 'fake' and genuine' were added to the facial expressions to make sure the participants knew what kind of expression (fake or genuine) they were looking at. Further, the Anger-to-neutral videos were included as that would allow to examine the difference between ratings obtained in the Joy-to-neutral and Anger-to-neutral, instead of using the one-sample t test with test value 3.00.

Two main predictions were put forward. The first prediction was that participants would evaluate the neutral expression at the end of the fake videos simply as neutral, as they knew the mental state of the actors at the start of the clips was not associated to any particular emotion. In this case there would be no response bias.

The second hypothesis followed a different reasoning. It was reported that some specific facial expressions (especially expressions of joy) are judged as more trustworthy (Krumhuber et al., 2007; Naumann, Vazire, Rentfrow, & Gosling 2009; Oosterhof & Todorov, 2009). Moreover, there is evidence suggesting that people who show genuine smiles are liked more (Young & Beier, 1977; Palmer & Simmons, 1995). Thus, it was reasonable to expect that when an actor is faking an expression and the observer knows that this is the case, the actor may be perceived 'negatively'. This negative attribution in turn (top-down) would affect the evaluation of the emotional signal (happy or angry), which means an overshoot with fake joy-sequences and an 'undershoot' with fake anger-sequences. The label "genuine" has a positive connotation and hence positive attributions may be made to the agent (again irrespective of the joy or anger emotion).

4.5.1 Method

4.5.1.1 Participants

Twenty-eight undergraduate Psychology students at Hull University (UK) took part in the experiment. All participants were recruited applying the same criteria as in Experiment 11 and 12.

After applying exclusion criteria (see data reduction below), the data of twenty-one participants were included in the analysis (age, M = 20.1 years, SD = 1.8 years; 15 females, 6 males). None of the participants took part in any of the other experiments.

4.5.1.2 Stimuli

Pictures depicting genuine and fake happy and angry facial expressions of the same four actors were used (Figure 4.2). Stimuli were created applying the same procedures as in Experiment 11 and 12. However, in the current experiment video-sequences consisted of both genuine and faked happy and angry expressions (Figure 4.9). Moreover, a red word (‘genuine’ or ‘fake’) was superimposed on the face midway up the nose. This massage appeared as soon as the video started but it was removed from the last neutral frame. The entire duration of the video-clips was 1170 ms (first frame 600 ms, last frame 300 ms and 9 interpolated frames 270 ms).

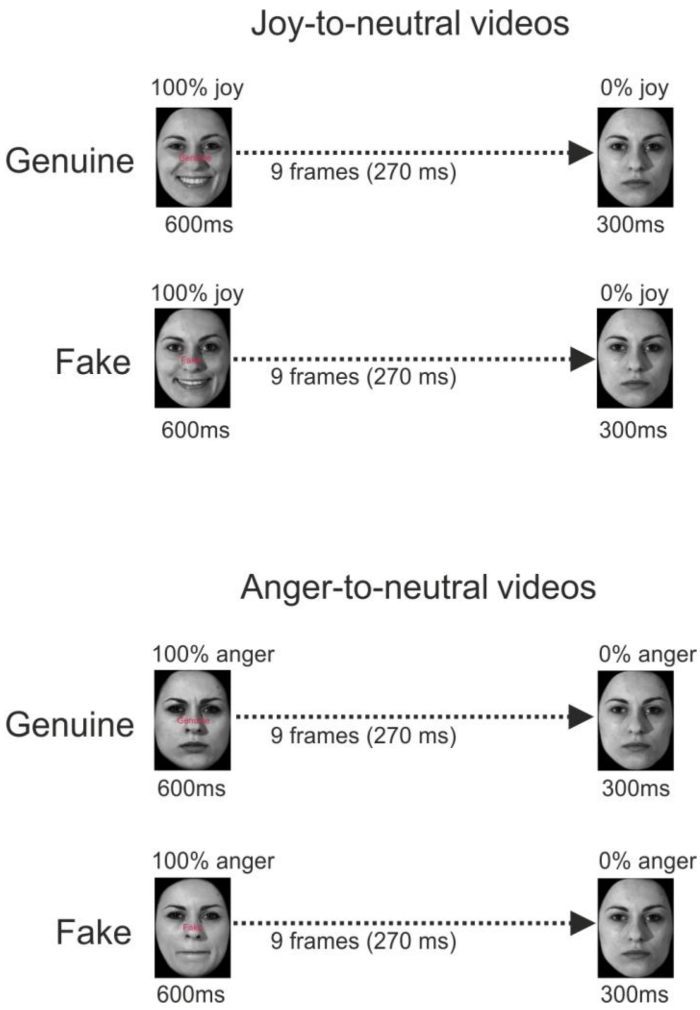

Figure 4.9. Schematic illustration of the stimulus presentations in Experiment 13. Genuine and fake Joy-to-neutral videos started with a facial expression of intense joy (600 ms), which gradually morphed into a neutral expression (top panel). Genuine and fake Anger-to-neutral videos started with a facial expression of intense anger (600 ms), which gradually morphed into a neutral expression (bottom panel).

4.5.1.3 Experimental procedure

As in Experiments 11 and 12, participants first made the calibration. They rated the neutral expression of each actor presented for 600 ms (8 trials: 4 actors X 2 repetitions). Next participants started the experimental session. In this phase participants were told that they would watch genuine and fake happy and angry expressions depicted by each actor. In order to make participants familiar with the differences between the two expressions the fake and genuine expressions of each actor were shown side by side (15 s). Participants were asked to watch carefully at the differences between the two types of expressions for each actor. After this ‘observation’ phase, the experiment started. Participants were instructed that half of the videos would depict posed, unfelt smiling and frowning expressions which are not linked to any emotional valence in particular. The other half of the videos would instead display genuine emotional expressions, thus linked to a positive (happy) or negative (angry) state of mind. First, 16 randomized practice trials, 4 actors X 2 perceptual histories (joy and anger) X 2 expressions (genuine and ‘fake’), were presented, followed by 64 randomized experimental trials (4 actors X 2 perceptual histories X 2 expressions X 4 repetitions). The duration of the entire experiment was 30 minutes.

4.5.1.4 Data reduction and analysis

Exclusion criteria were the same as for Experiment 11 and 12. Trials in which the RTs fell below 250 ms or above 3000 ms were removed (8.1% of total number of trials).

Participants were excluded if more than 25% of the RT values fell outside the above range (n = 4) and if they exceeded the $\pm$2.5 SD rule to the mean difference ratings (n = 3). The calibration scores for the neutral expression of each actor are illustrated in Figure 4.10. All statistical analyses were performed on the calibrated scores.

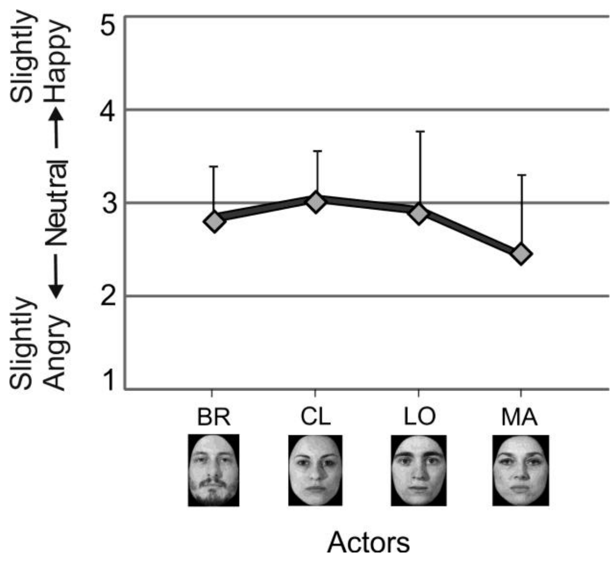

Figure 4.10. Calibration scores in Experiment 13. Ratings on the 5-point scale (vertical axis) for the neutral expression presented in isolation. Scores are shown for each of the four actors Error bars indicate +1SD.

4.5.2 Results

The results are shown in Figure 4.11. A 2X2 repeated measures ANOVA with Perceptual history (joy vs. anger) and Expression (genuine vs. fake) as within subjects factors showed a strong significant main effect for Perceptual history (F(1, 20) = 41.44, p < .0001, η_p^2 = .674), which reflected the usual overshoot effect: the expression in the last neutral frame was perceived as more angry in the joy videos (M = 2.74, SD = 0.55) than in the anger videos (M = 3.43, SD = 0.45). There was no main effect for Expression (F(1, 20) = 1.13, p = .301, η_p^2 = .053). The two-way interaction History by Expression was also not significant (F(1, 20) = .222, p = .643, η_p^2 = .011). In the genuine condition neutral faces were perceived as more angry in the Joy-to-neutral videos (M = 2.70, SD = .56) than in the Anger-to-neutral videos (M = 3.42, SD = .37), ($t(20) = -6.21$, $p < .0001$), and each differed significantly from 3.00 (joy: $t(20) = -2.42$, $p = .025$; anger: $t(20) = 5.16$, $p < .0001$). In the fake condition, neutral faces were perceived as more angry in the Joy-to-neutral videos (M = 2.77, SD = .54) than in the Anger-to-neutral videos (M = 3.44, SD = .52; $t(20) = -5.34$, $p < .0001$). However, only the evaluation of the neutral expression at the end of the Anger-to-neutral videos differed significantly from 3.00 ($t(20) = 3.85$, $p = .001$), while it approached significance for the Joy-to-neutral sequences ($t(20) = -1.94$, $p = .067$).

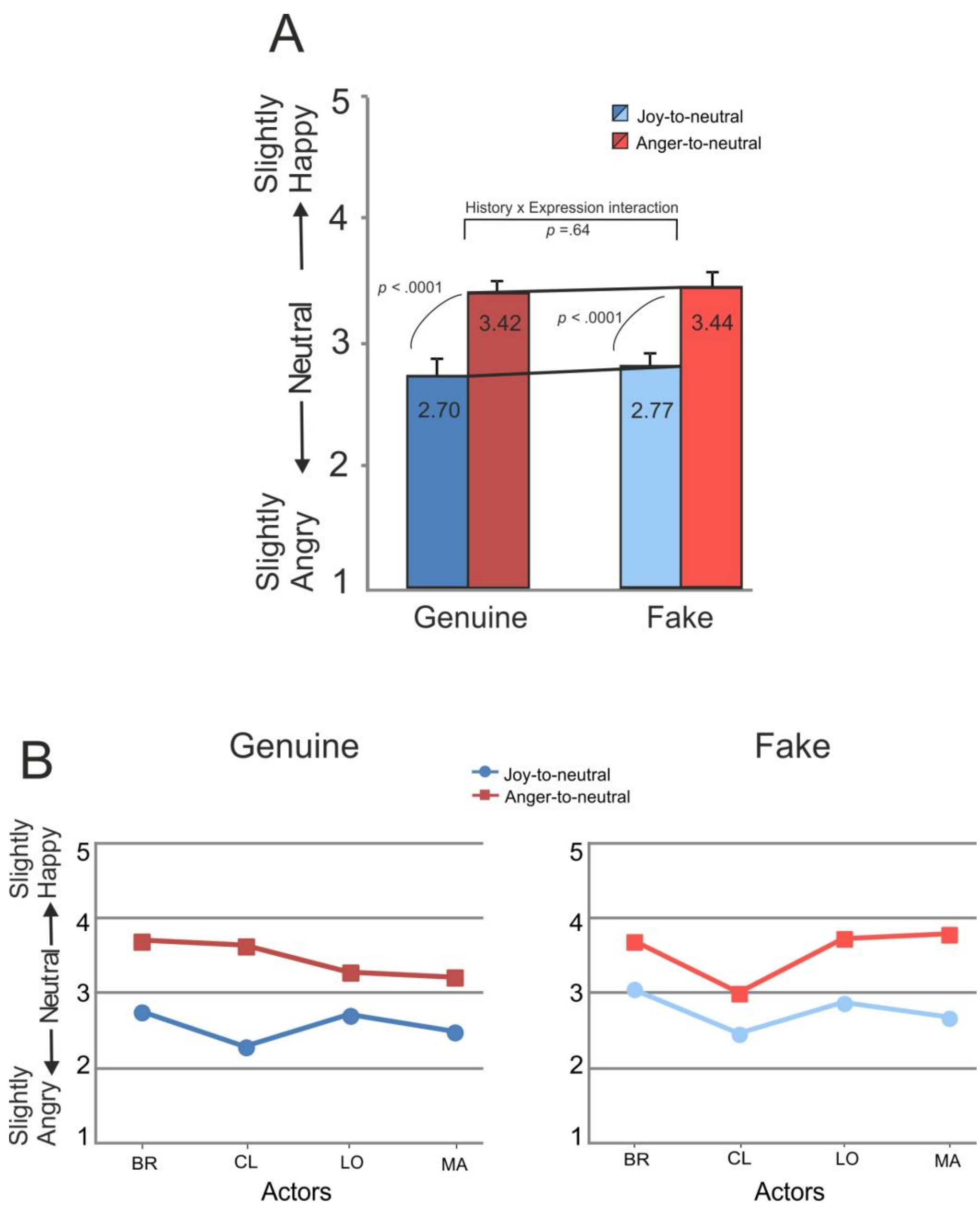

Figure 4.11. Results of Experiment 13. (A) Scores for the neutral expressions at the end of the genuine Joy-to-neutral and Anger-to-neutral sequences (horizontal axis). Error bars indicate SEM. (B) Ratings for each of the four actors with genuine Joy-to-neutral and Anger-to-neutral videos (left side) and with fake Joy-to-neutral and Anger-to-neutral videos (right side) are shown to illustrate the consistency across actors.

4.5.3 Discussion of Experiment 13 & General Discussion of Chapter 4

Experiment 13 was conducted to examine whether the overshoot would be affected by videos of the same actors expressing genuine/fake smiling and frowning expressions, while participants were informed about whether the expressions were fake or genuine. This manipulation was done to make sure the participants knew which expression was fake and which was genuine.

An overshoot bias was found with both genuine and fake Joy-to-neutral and Anger-to-neutral videos. Moreover, the amount of overshoot did not differ between the two conditions. Assuming that the observers took account of the information provided, then our hypotheses about the influences of the fake emotions have to be rejected.

One possible explanation for the lack of a difference between the fake and genuine conditions is that participants simply did not pay attention to the labels " genuine" and "fake", as the labels were in no way necessary to perform the task (i.e. evaluate the last facial expressions). Some participants indeed indicated in the debriefing that they basically had ignored the labels. However, if it is assumed that most participants did take the labels into account, then another possibility for the lack of a difference between fake and genuine is that the fake expressions did not evoke an 'experience' of fakeness in the participant. The SIMS model (Niedenthal et al., 2010) claims that such an experience is crucial in order to perceive the expression as fake. Theories on embodiment of facial expressions put forward by e.g. Gallese (Gallese, 2007) and others (Carr et al., 2003; Goldman, 2006; Jabbi et al., 2007; Niedenthal, 2007; Pitcher et al., 2008, Wicker et al., 2003) also claim that such an 'experience' is essential for the understanding of the observed emotion. The reason why the experience of fakeness may not have been induced in the observers may be that the distinctive temporal components of faked expressions were lacking. That is, according to the literature one of the most distinctive characteristic of fake smiles is that they posses more irregular actions, phases, pauses, and stepwise intensity changes than spontaneous smiles which have slower onsets and offsets (from the peak of the expression to the neutral expression or baseline) durations (Ekman & Friesen, 1982; Hess & Kleck, 1990; Krumhuber et al., 2007; Schmidt et al., 2006; Weiss, Blum, & Gleberman, 1987). As in our experiment the offset was not manipulated and the transition to neutral was smooth, which could have rendered the fake smiles and frowns less realistic, the visual input provided by the video-clips may have activated the same motor response in the observer for the fake and genuine videos, thus resulting in the overshoot bias in both the conditions.

Therefore, in future experiments using this paradigm, a recognition task for the joy and anger offsets is needed, especially because the current stimuli were not created using the FACS system (Ekman & Friesen, 1976). This would clarify whether the offset of the expressions was appropriate or whether the temporal and dynamic components of the fake expressions should be taken into account to make it more realistic. Further, real-time recordings of the offset of genuine and fake expressions could be used, in line with the reviewed literature on deceptive facial expressions. If the role of the dynamic component of the video-clips would turn out to be crucial, then this would strongly boost the claim that embodied simulation plays a crucial role in understanding the meaning of facial expressions (Goldman & Sripada 2005; Gallese, 2006a, 2006b). Importantly, it would also suggest that ‘motor resonance’ may serve emotional anticipation, which allows people to anticipate how the emotional state of the other person would develop in the immediate future (Palumbo & Jellema, under review). This mechanism is extremely relevant to navigate in the social world and face the complexity of social interaction and interpersonal communication in very different environments and domains.

CHAPTER 5

General discussion

Humans make extensive use of socially relevant information in everyday interpersonal relations. Successful social interactions mainly depend on the ability to explain those signals in terms of epistemic mental states (Baron-Cohen, 1995). Understanding and anticipating intentions, beliefs, emotions, desires and motivations, which govern other people's behaviour, is a key mechanism of social cognition.

Mental state attributions may derive from the visual analysis of a variety of social signals. This largely bottom-up perceptual evaluation provides a mechanistic description of social cues, which is seen as an essential step towards the comprehension of others' mental/emotional states (Jellema & Perrett, 2002a, 2002b; 2005; Balkemore & Decety, 2001). However, the process instantiated by the detection and description of social cues leading up to 'mentalizing' involve also a feedback (top-down) route. Interestingly, recent work has suggested that interpretations of other people's intentions based on explicit information can exert a top-down influence, which is strong enough to affect basic low-level visual processes, such as adaptation, which were typically considered to be immune from such top-down processes (Teufel et al., 2009). This gave rise to the concept of 'perceptual mentalizing' (Teufel et al., 2010), which emphasizes the bidirectional relation between perceptual analysis and higher order cognitive function. Importantly, it has been reported that such top-down influences may also proceed in an implicit, automatic manner, without the use of explicit information (Hudson et al., 2009).

The general aim of this thesis was to investigate how the perceptual processing of social cues may interact with mental attributions, which are generated implicitly and involuntarily on the basis of the immediately preceding perceptual history of those social cues. The first objective was to assess whether involuntary and automatic mental/emotional attributions, during observation of dynamic facial expressions, may lead to an anticipation of the most likely future emotional state of the agent, which in turn (top-down) would influence the evaluation of those expressions. Specifically, the aim was to clarify the nature of the so called ‘overshoot effect’, a perceptual distortion occurring in the evaluation of happy and angry facial expression offsets (Jellema et al., 2011). In Jellema et al. (2011) video-clips of happy and angry facial expressions that gradually morphed into a neutral expression were presented. Participants evaluated this last neutral expression by using a 5-point scale ranging from slightly angry via neutral to slightly happy. Results showed that the last neutral expression in the joy offsets was judged as slightly angry, while the identical neutral expression at the end of the anger offsets was rated as slightly happy (overshoot bias). Thus, the evaluation of the last neutral frame of the videos was affected by the specific perceptual history (Joy-to-neutral or Anger-to-neutral). Importantly, when the video-clips where used as task-irrelevant distracters in an affective priming paradigm (Jellema et al., 2011), it was shown that the response bias could not have been caused by a cognitive response strategy (i.e. selection of the emotion opposite to the start-emotion). Rather, the affective priming study suggested that the Joy-to-neutral and Anger-to-neutral sequences induced negative and positive affects in the observer, respectively. Thus, a speculative interpretation of these findings was that the perceptual history witnessed immediately prior the last neutral to-be-evaluated frame may have triggered an anticipation of the forthcoming affective state of mind of the agent. This hypothetical mechanism was referred to as ‘*emotional anticipation’* (cf. Goldman, 2006), which in turn could have biased the perception of the neutral expression in the direction of the anticipated state of mind. In everyday life people continuously need to understand and predict the reasons underlying others’ actions in order to react appropriately. According to some theoretical positions the valence of stimuli, e.g. 'good' or 'bad', is detected at early stages and it triggers the motivational systems of approach and avoidance (Bargh, 1997; Lang, Bradley, & Cuthbert, 1990). From an evolutionary perspective, approach and avoidance responses are adaptive in that they contribute to the survival of the individual. Adopting this perspective, a decrease of the intensity of a smile can be associated to a dangerous signal which can activate an avoidance reaction; while a decrease of the intensity in the anger expression may activate an approach response. The ability to anticipate changes in someone else's affective state is fundamental to adjust one's own behavior. In some cases it can also serve manipulatory purposes (Weiner & Handel, 1985; Petty & Cacioppo, 1986). In Jellema et al. (2011) no attempts were made to clarify the role of emotional anticipation in explaining the overshoot response bias and to rule out other plausible interpretations. In fact, a variety of low-level visual processes may offer good explanations for the presence of this perceptual response bias. These processes include contrast/context effects (Russell & Fehr, 1987; Suzuki & Cavanagh, 1998; Tanaka-Matsumi et al., 1995; Thayer, 1980), adaptation (Fox & Barton, 2007; Hsu & Young, 2004; Leopold et al., 2001; Rutherford et al., 2008; Webster et al., 2004), extrapolation of the geometrical features in the face (U shape in a smile) and RM on an underlying positive-to-negative valence dimension (Yoshikawa & Sato, 2008).

The focus of the studies presented in Chapter 2 was to disentangle the contributions of these basic bottom-up perceptual processes and the contribution of top-down emotional anticipation, to the generation of the overshoot bias during the observation of relatively fast offsets of joy and anger facial expressions. To this end, six experiments were conducted with typically-developed (TD) individuals.

In Experiment 1 the end-point of the joy and anger videos was varied (between neutral, 10% joy and10% anger expressions). This manipulation allowed assessing the presence of the overshoot bias also when the last expression belonged to the same emotional category as that of the starting expression. It was found that in these cases the overshoot bias did occur, overcoming a 20% difference in emotional intensity.

In Experiment 2 the identity of the actors was changed immediately before displaying the last neutral frame. Two versions of this Identity-change manipulation were applied. In the instant Identity-change condition the identity of the actor was changed in the very last neutral frame of the videos, whereas in the smooth version the identity change was realized through a gradual morphing between the two identities. In both cases the degree of emotional contrast between the first and last frames was unaltered. The overshoot bias was completely removed in the Instant identity-change condition and was significantly reduced in the smooth identity-change (the latter result was probably due to participants not detecting the identity transition due to morphing between identities).

The overall result suggested that sequential contrast/context effects (see Thayer, 1980) probably did not play a major role in bringing about the response bias. Furthermore, an explanation in terms of an extrapolation on an underlying positive-to-negative valence dimension (RM) was unlikely as this underlying dimension was not affected by the identity-change manipulation. Experiment 2 seems to also rule out the role of visual adaptation as the change in identity would not have removed any aftereffects (Fox & Barton, 2007; Ellamil et al., 2008).

In Experiment 3 a Loop condition was introduced, in which the first frame depicted a neutral expression, which subsequently morphed into an expression of either joy or anger, after which it morphed back into the same neutral expression.

The results again did not support a contribution from sequential contrast effects and RM. The last neutral frame of the Neutral-to-joy-to-neutral sequence was judged as slightly happy, while the identical last neutral frame of the Neutral-to-anger-to-neutral was judged as slightly angry (i.e. an ‘undershoot’ response bias). A likely explanation for this finding is that the brief smile in the Neutral-to-joy-to-neutral videos indicated a positive social signal, and the brief frown depicted in the Neutral-to-anger-to-neutral sequence a negative social signal. The video-clips appeared therefore to display a daily social interaction where the observer may have attributed a positive disposition to the agent in the Neutral-to-joy-to-neutral condition and a negative disposition in the Neutral-to-anger-to-neutral condition. The finding argued against sequential contrast effects because in case of a sequential contrast effect one would expect the last neutral expression to be evaluated as neutral and not obtain a response bias. It also excluded RM effects on the underlying positive-negative valence dimension as an explanation as this should have resulted in an overshoot bias.

A possible contribution of RM effects on dynamic facial expressions was further discredited by Experiment 4. A manipulation of the speed of the video-clips did not affect the overshoot bias, contrary to the study by Yoshikawa & Sato (2008) who reported that RM effects induced by their video-clips depicting neutral-to-emotion dynamic expressions were modulated by the speed of the videos.

In Experiment 5 the insertion of a 400 ms long mask immediately before the last, neutral expression did not affect the overshoot bias. This indicated that the overshoot bias did not depend on the uninterrupted flow of motion present in the video clip (cf. Enns et al., 2010).

Finally, the contribution of adaptation was specifically examined in Experiment 6. This experiment showed that the overshoot was not modulated by doubling or even tripling the duration of the first frame (from 200 ms to 600 ms) of the videos (Jellema et al., 2011). Taken together the findings reported in Chapter 2 suggest that the overshoot bias cannot be considered to be a result of bottom-up visual basic processes. Alternatively, the direction of the bias is congruent with the emotional anticipation hypothesis, which would explain the overshoot as the consequence of top-down attributional processes.

One could ask why an agent displaying a joyful facial expression that subsequently (gradually) changes into a neutral expression would be perceived as adopting a ‘negative’ state of mind. Why is the agent not seen as adopting simply a neutral state of mind? There are at least two lines of research that indicate that emotion-to-neutral sequences induce an emotional valence in the observer (which in turn may be attributed to the agent). The first line of research involves affective priming paradigms (Jellema et al., 2009). In these studies happy/angry-to-neutral sequences were used as distractors, while positive or negative words superimposed on the last (neutral) frame of the videos are used as targets. It was found that negative words were faster detected than positive words in the Joy-to-neutral condition, whereas positive words were faster detected in Anger-to-neutral condition. This strongly suggests that the sequences evoke specific affective valences in the observers.

The second line of research involves imaging of brain activity during presention of joy and anger offsets (Mühlberger et al., 2011). This study suggested that the joy offset was perceived as a threat signal, inducing an avoidance reaction and the attribution of a negative emotional state to the agent. In contrast, the anger-to-neutral offset could have assumed a reward valence, thus generating approach-like behaviour and the attribution of a positive state of mind to the agent. The findings presented in Chapter 2 allow to reject explanations for the overshoot phenomenon in terms of basic perceptual processes and are compatible with the emotional anticipation hypothesis.

In Chapter 3 four experiments aimed at testing the same emotional anticipation tasks on individuals with Asperger's syndrome (AS) and a control group of TD individuals, matched for age, gender and intellectual abilities. Individuals with AS, although they manifest many intact cognitive functions, show specific difficulties in spontaneous emotion reading and in the intuitive understanding of other people's behaviour (Jellema et al., 2009; Kuchine et al., 2011; Kuzmanovic et al., 2011; Senju et al., 2009). Therefore, the study of individuals with AS might shed light on the explanation of the overshoot bias in terms of emotional anticipation. If individuals with AS are indeed compromised in emotional anticipation than one would expect not to find a perceptual distortion with this clinical population. However, previous studies suggested that anticipatory skills concerning other people's behaviour may be present in AS individuals, although it was suggested that this could be derived from a purely mechanistic analysis of social signals (Hudson et al., 2011), without involving the agent's underlying psychological states.

In Experiment 7, AS individuals showed the same response bias as the one found for the TD participants (Experiment 1, Chapter 2). The last expression of the Joy and Anger videos was evaluated as slightly angry and slightly happy, respectively. This occurred even when the last frame belonged to the same emotion category as the start emotion. This finding left open two possibilities: (1) the individuals with AS are not impaired in emotional anticipation, and (2) they are impaired in emotional anticipation but apply compensatory mechanisms. Such compensatory mechanisms could in principle have relied on low-level visual processes like contrast effects, RM on facial features and RM on positive-negative valance dimension. In order to test whether (any of) those mechanisms were involved in producing the perceptual bias, three manipulations adopted in Chapter 2 were also applied in the studies of Chapter 3: instant Identity change (IC, Experiment 8), insertion of a Mask (Experiment 9) and the Loop sequence (Experiment 10).

In Experiment 8 remarkably, the Identity change manipulation did not remove the overshoot bias in the AS group. In the control TD group, the Identity change did remove the overshoot bias, similarly to the findings of Experiment 2. Importantly, the presence of the response bias in this condition in the AS group was not related to a failure to detect the Identity-change, as the same extent of overshoot was found in trials where the gender of the two identities was the same compared to trials which involved a gender shift. The results obtained with AS individuals suggested that their overshoot bias was probably not generated by emotional anticipation (as emotional anticipation would have predicted no response bias in the Identity change condition), but by alternative mechanisms, such as contrast effects or extrapolation (RM).

In Experiment 9 TD and AS participants both showed an overshoot effect when a 400 ms mask, inserted just before the last neutral frame was displayed, interrupted the flow of the motion. This suggested that the hypothesized low-level perceptual mechanism causing the overshoot response bias in the AS group was not dependent on the continuous flow of motion of the dynamic sequence (cf. Enns et al., 2010). Disruption of the flow did not affect the overshoot. Experiment 10 tested whether contrast and RM effects might have been responsible for the generation of the response bias in the AS group by using the Loop condition. If contrast effects were responsible then the prediction was that the AS group would not produce any bias. On the other hand, RM would predict an overshoot. Interestingly, AS participants did not show significant biases in the Loop condition, i.e. the last neutral frame was judged as nearly neutral. The lack of bias in the Loop sequence, and the presence of the overshoot in the No loop sequence, suggested that in AS individuals the perceptual distortion might have been due to contrast effects. The results of the TD individuals confirmed the findings of Experiment 3, by revealing an ‘undershoot’ effect in the Loop condition.

The Identity change (Experiment 8) and the Loop (Experiment 10) conditions were crucial tasks as they discriminated between TD and AS participants. In the Identity change condition TD participants could not anticipate the emotional state of agent B as all they had witnessed was the immediate perceptual history pertaining to agent A. Thus, TD participants evaluated agent’s B expression as neutral. However, in AS individuals, the overshoot bias occurred even if the information needed to anticipate the emotional state of agent B was not available. This suggests that the evaluations of AS participants were not caused by emotional anticipation. The Loop task clarified that the perceptual distortion found in AS individuals was likely due to visual contrast effects.

Overall the studies reported in Chapter 3 suggested that AS participants based their judgments on the visual perceptual characteristics of the displayed dynamic facial expressions, rather than on the associated emotional meaning, which was automatically captured by TD participants (Experiments 8 and 10). In line with previous evidence of compensatory mechanisms for action anticipation in AS individuals (Hudson et al., 2009), it is proposed that participants with AS solved the emotional anticipation task by applying mechanistic, rule-based strategies, which are typically adopted by AS individuals to understand non-agentive events (Jellema & Perrett, 2002a, 2002b; 2007). Such strategies, based on refined cognitive skills, might compensate for their impairment in emotion-reading and possibly for the lack of social understanding.

In Chapter 4, another attempt was made to establish the role of emotional anticipation in causing the overshoot bias in TD individuals involving deceptive facial expressions of emotions. The rationale of this study was that if the overshoot effect was indeed generated by the ability to involuntarily anticipate how the emotional state of the agent would develop in the immediate future on the basis of the immediate perceptual history, then the introduction of a discrepancy between the facial expression and the underlying emotional state of mind should affect the response bias. That is, if the perceptual history include untruly emotional expressions. To this end a new set of stimuli was created in collaboration with professional actors who displayed genuine and fake happy and angry facial expressions. Once the presence of the typical overshoot response bias with the genuine expressions depicted by the new actors was confirmed (Experiment 11), two additional experiments were conducted to test whether the 'fakeness' could disrupt the response bias. According to the emotional anticipation hypothesis, a reduction of the overshoot bias should be found for the fake Joy-to-neutral video-clips as compared to the genuine ones, because the agent's emotional state did not start as really happy. It was therefore hypothesized that participants would judge the neutral expressions at the end of the fake video-clips as (nearly) neutral. In Experiment 12 the fake and genuine Joy-to-neutral sequences produced equal overshoot biases. However, this perceptual bias was less strong in both conditions. One possibility is that the lack of contrast with the Anger-to-neutral sequences may have affected the results, in that it made of an overshoot bias more difficult to verify. As it was possible that the participants simply had not recognized that some of the expressions were fake, in Experiment 13, where Anger-to-neutral vides were also employed, labels superimposed on the faces indicated whether the expression was fake or genuine.

This also allowed examining the role of explicit knowledge on the evaluations of those expressions. A similar overshoot bias was found in the genuine and fake conditions (for both joy and anger perceptual histories), meaning that apparently, explicit knowledge did not influence their perceptual evaluations. However, it cannot be excluded that participants simply did not take account of the information provided. The fake expressions were realized taking care of the differences in the facial features between fake and genuine expressions (asymmetry of the mouth, lack of Duchenne markers). It is possible, however, that these differences were too subtle, also because of a lack of proper dynamicity (Ekman et al., 1982; Hess & Kleck, 1990; Krumhuber et al., 2007; Schmidt et al., 2006) in the fake video-clips.

However, the decision to present the video-clips with the same temporal and dynamic properties as in the previous studies (Chapters 2 and 3) was motivated by the necessity to exclude explanations for the overshoot based on purely basic visual processes.

5.1 Emotional anticipation as an implicit mechanism of social understanding

The results reported in this thesis suggest that TD individuals are equipped with an automatic and mandatory mechanism, which allows them to predict and anticipate how the emotional state of other individuals may change according to immediately preceding events. This mechanism is extremely important to navigate in the social world. People who lack the ability to generate expectations about how another person may react when confronted with a particular emotional event can find it extremely difficult to interact socially. This would include difficulties in cooperating with others, and in sharing intentions and motivations. Individuals with AS can process social cues, and describe social situations, in mere mechanistic terms without necessarily grasping the mental states associated to others' actions or the emotional meaning beyond their facial expressions (Senju et al., 2009; Jellema et al., 2009). This may suggest that they can have expectations about how a certain situation may develop in the immediate future, but they would find it difficult to understand its psychological implications. As such they would lack the ability to react appropriately and promptly when it is needed. Even though individuals with AS can use their cognitive skills to compensate social impairments, on the other hand such compensation may be not sufficient to repair their social fragility.

In the last three decades the scientific debate focused mainly on assessing how social understanding is realized. Traditionally, having a Theory of Mind (Baron Cohen, 1995) refers to the explicit or implicit adoption (Fodor, 1968; see also Stich & Nichols, 1992) of a flexible and updated theoretical module, or 'folk psychology', of stored symbolic representations and concepts that enables inferences about covert mental states from the observation of overt behaviour. One of the challenging questions is whether interpretations and predictions about others' mental states underlying social behaviour is instantiated by high-level cognitive functions (Csibra, 2007; Jacob, 2008), generating representations of the psychological meaning of actions. Alternatively, it has been proposed that people can grasp, at least for observed simple actions that are part of one's own experiential repertoire, the associated mental contents by embodied simulation (e.g. Gallese, 2007). According to Gallese's account, perceived actions, emotions and sensations of others are mapped onto the observer's motor system, thus triggering internal, prelinguistic representations of the body states that correspond to those actions. This would result in an '*as if'* state in the observer, as if the observer was performing a similar action or feeling a similar emotion as the perceived agent. This is a less sophisticated, mandatory and reflexive pathway, based on a direct matching module served by the mirroring mechanism (Gallese et al., 1996; Gallese & Goldman, 1998; Rizzolatti et al., 2001; Rizzolatti & Craighero, 2004).

The concept of emotional anticipation fits within models of embodied cognition of facial expression (Carr et al., 2003; Goldman, 2006; Jabbi et al., 2007; Niedenthal, 2007; Niedenthal et al., 2010; Pitcher et al., 2008, Wicker et al., 2003), which emphasize the prominent role played by one's own motor experience while observing emotional facial expressions.

The results presented in this thesis suggest that emotional anticipation may not involve the mediation of a ToM, based on mental state concepts or reasoning. Rather, it is speculated that the results may be better explained in terms of a low-level mindreading mechanism (Goldman, 2006; Goldman & Sripada, 2005), underpinned by embodied simulation (Gallese, 2007). Specifically, a simulation of the agent's dynamic facial expression activate motor schema that will give rise to the emotional state the observer would arrive at if the observer would experience the same dynamic facial expression. This may have triggered an internal anticipation of how the emotional state of the agent would have developed in the immediate future, which in turn influenced the perception of the facial expressions in question.

This is in line with the perceptual mentalizing model proposed by Teufel et al. (2010), the anticipation of how the emotional state of the agent would have developed if the video-clip would have continued beyond the stopping point influences the perceptual evaluation of the last neutral frame of the sequence. However, in contrast to Teufel's model, where explicit information offered to participants influenced perception, in the model presented in this thesis the information was provided implicitly within the immediate perceptual history (the video-clips) and emotional anticipation was implicitly and automatically generated by this perceptual history. These results are corroborated by previous studies that showed how the perceptual history implicitly provides information about the agent's behavioural intentions, which is 'picked up' by the observer and influences perception (Hudson et al., 2009, Hudson & Jellema, 2011).

Thus, here the distinction between top-down inferential processes, which includes a mentalistic analysis of social events and bottom-up perceptual processes that lead to a mechanistic description of the social event is revisited, taking into consideration the possible involvement of the mirror mechanisms. The study presented in Chapter 2 opens up the possibility that embodied simulation may serve as an intermediate step between a purely mechanistic visual analysis of the social stimuli and high-level theorizing about others' emotional states based on the perceptual analysis of those social cues, contextual information and knowledge (Palumbo and Jellema, under review).

The data reported in Chapter 3 from individuals with AS supports the hypothesis of emotional anticipation with TD individuals and suggest the use of compensatory, rule-based strategies adopted by AS people to cope with social situations. These findings are also in line with the 'broken mirror' hypothesis, which suggests that an impairment of the mirroring mechanism causes a lack of emphatic reactions, which then affects the judgements of facial expressions (Williams et al., 2001; Dapretto et al., 2006; Wang et al., 2004). However, direct evidence supporting the hypothesis of a direct matching pathway underpinning emotional anticipation (and its dysfunction) is not yet provided and further investigations are necessary.

The study using fake facial expressions (Chapter 4) tried to provide additional insights about the mechanism of emotional anticipation and focused on one issue in particular: What if the emotional expression and the emotional state of mind are not congruent? Normally, when a person smiles she is happy and when she frowns she is angry. In these cases the facial expression and the emotional state of mind go hand in hand. However, in the case of faked smiles and faked frowns there is a discrepancy between the facial expression and state of mind. In Chapter 4 the emotional anticipation hypothesis was tested by exploiting this discrepancy. Does an overshoot response bias occur if the emotional state of mind did not change much, while the expression did change? Emotional anticipation would predict a reduced or absent overshoot bias. Do people understand the ‘why’ of an expression by a bottom-up motor resonance? The idea is that the visual information about the observed expression is first analyzed by the superior temporal sulcus (STS) and then passed on to the parietal mirror neuron areas and from there to the mirror areas in the premotor cortex. Thus, the meaning beyond the physical features of the expression may be processed within the motor system.

In Chapter 4 the visual information contained in the fake and genuine offsets was very similar. If the perceptual evaluation of the video-clips resulted in a simulation of those expressions, then it is possible that the outcome of this simulation mechanism was the same for both genuine and fake expressions. In other words, the underpinning emotions/intentions were not simulated correctly because of a lack of adequate visual input to the observer’s motor system. This, in turn, could have affected emotional anticipation and generated the overshoot bias in both the fake and genuine videos. It is important to note that in Experiment 3 participants were aware of the kind (fake vs. genuine) of video-clips they were watching, but it seems that explicit knowledge did not modulate the responses.

The action '*reconstruction model*' (Csibra, 2007) suggests that a hypothesized goal of an action is produced prior the generation of the action's template in the observer's motor system. Thus, "action understanding may proceed, rather than follow, action mirroring" (Csibra, 2007, p. 443). However, the results of the 'fake' experiments suggest that explicit knowledge did not affect the evaluations. Possibly, the motor simulations induced by the fake and genuine expressions were simply too similar, making it impossible for the observer's mirror mechanism to discriminate between them. Therefore, the overshoot bias reported here, being the product of involuntary anticipation based on motor simulation was the same in both conditions, despite explicit information indicating otherwise. This would mean that motor resonance is an automatic and mandatory mechanism for producing an involuntary anticipation of the affective states of other people, which overrides explicit knowledge. This is in line with research carried out by Maringer et al. (2011), who reported that only when the automatic mimicking (i.e. simulation) of the facial expression was inhibited (by holding a pen in the mouth) the context, stereotypes and beliefs were taken into account in the evaluation of the meaning of smiles. However, at the moment this is just a speculative interpretation of the results presented in this thesis.

The current debate in Theory of Mind leaves open some fundamental questions about the role of the direct matching account in implicit mentalizing processes. (1) Does a mirroring mechanism provide a direct tool to discriminate between different kinds of mental states when these are displayed by similar behavioural acts? (2) As mental states (e.g. intentions, beliefs and motivations) tend to be more complex than emotional states, is it possible that the kind of role played by simulation and mentalizing routes may depend on this complexity? (3) Can a simulation account be sufficient to explain the ability to anticipate what is beyond people's emotional facial signals? (4) What is the role of motor representations and embodied simulation in social understanding and in the involuntary attribution of emotional/mental states to others? Does embodied simulation form a crucial stage in the formation and the use of affective state concepts?

5.2 Outlook for future research

Even though the nature of the experiments presented in this thesis did not allow to provide a definitive confirmation of a top-down account involving motor simulation as explanation for the overshoot response bias, the experiments did manage to exclude alternative candidates involving basic visual processes (Chapter 2). This conclusion was supported by the differences between TD individuals and the AS group found in the Identity change and Loop tasks (Chapter 3). Future investigations could elaborate on the studies presented in this thesis in the following ways.

(1) The emotional anticipation hypothesis could be tested with additional facial expressions. In the current experiments, angry and happy expressions were chosen because both are approach-oriented (cf. Davidson & Hughdahl, 1995) and are fairly similar in terms of arousal, but occupy opposite positions on the valance dimension (Russel, 1980; Russel & Bullock, 1985). However, in principle the ability to attribute an affective state to the agent and form expectations about how that state may change over time based on the immediate perceptual history and/or contextual information can also be applied to other emotional displays, which include different facial expressions but also emotional body postures.

(2) It is known from an affective priming study (Jellema et al., 2011) that the neutral frame at the end of Joy-to-neutral clips is associated with a negative affective valence, and likewise that the neutral frame at the end of Anger-to-neutral clips is associated with a positive affective valence. This is suggestive of embodied simulation, which can be further assessed using EMG. Specifically, it would be relevant to test whether observation of the Joy-to-neutral video clips would induce an initial activity in the EMG of the observer's 'smiling' muscles (Zygomaticus Major), followed by activity in the 'frowning' muscles (Corrugator Supercilii) at the end of the clip during (and following) presentation of the neutral expression. This would be surprising as in these video-clips there is no angry facial expression at all. Similarly, it would be relevant to see whether observation of the Anger-to-neutral video clips would induce an initial activity in the EMG of the observer's 'frowning' muscles, followed by activity in the 'smiling' muscles at the end of the clip. Again, that would be surprising as in these video-clips there is no happy facial expression at all. If these hypotheses are confirmed then that would indicate embodied simulation as the underlying mechanism for the perceptual overshoot bias. Experiment 14 (Appendix) made a start with this line of investigation, but more participants (both TD and AS) will have to be tested before firm conclusions can be drawn.

(3) The video-clips for the fake condition lacked the appropriate dynamicity that makes an expression being implicitly recognized as faked (Ekman et al., 1982; Hess & Kleck, 1990; Krumhuber et al., 2007; Schmidt et al., 2006). However, at the present no information is provided in the literature about the temporal dynamicity of the offsets of faked angry expression and further investigations are necessary.

Moreover, in the study presented in Chapter 4 only four actors were used thus introducing more variability and familiarity effects, as compared to the studies in Chapters 2 and 3 where eight actors from the Ekman and Friesen (1976) validated set were chosen. Ideally, a large set of identities should be used.

(4) At the present the SIMS model (Niedenthal et al., 2010) exclusively applies to smiling expressions. Nothing is known about negative facial displays. The study reported in Chapter 4 represents the first attempt to investigate pretended angry expressions. If the simulation model can account for the immediate understanding of the meaning of smiles, then this should be also the case for negative facial affects. Further experiments could incorporate real-time video-recordings.

(5) A new study may also involve EMG responses of the observer's facial muscles while watching genuine and fake smiles. This would clarify whether different facial units would be mimicked during the observation of a fake smile as compared to a genuine one. In the case of 'direct' matching motor resonance the activation of congruent facial muscles is expected. Use of the EMG technique during the 'emotional anticipation' task with fake and genuine Joy-to-neutral videos would allow to test the role played by 'direct matching' in anticipating the emotional state of mind the agent is going to have in the immediate future.

(6) Further, more compelling, evidence about the possible involvement of mirror mechanism in emotional anticipation might be obtained by using the current video-clips in fMRI studies. Similar anger and happy offset stimuli as used in this thesis, but depicted by schematic figures (made with Poser software), were presented by Mühlberger et al. (2011) in an fMRI study. They reported that the offset

of happy and the onset of angry expressions both signal a threat and lead to an avoidance reaction, as they activated the typical avoidance areas: the lateral orbitofrontal cortex, the left amygdala and the right insula. The onset of happy and offset of angry expressions signal a reward and induce an approach reaction, as they activated the typical approach area: left dorsal striatum. However, these authors did not focus their analysis on the mirror neurons areas. Such a focus could be applied in future fMRI studies to further test the motor simulation hypothesis.

References

Abel, M. H. (2002). The elusive nature of smiling. In M. H. Abel (Ed.) *An empirical reflection on the smile* (pp. 1-13). Lewiston, NY: Edwin Mellen Press.

Abell, F., Happe, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development, 15*, 1-16.

Adams, R. B., & Kleck, R. E. (2003). Perceived gaze direction and the processing of facial displays of emotion. *Psychological Science, 14*, 644-647.

Adolphs, R. (2009). The Social Brain: Neural Basis of Social Knowledge. *Annual Review of Psychology, 60*, 693-716.

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience, 4*, 165-178.

Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology, 12*, 169-177.

Adolphs, R. (2000). What emotions really are. *American Journal of Psychology 113*, 472-478.

Adolphs, R. (1999). Social cognition and the human brain. *Trends in Cognitive Sciences, 3*, 469-479.

Adolphs, R., Sears, L., & Piven, J. (2001). Abnormal processing of social information from faces in autism. *Journal of Cognitive Neuroscience*, *13*, 232-240.

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired Recognition of Emotion in Facial Expressions Following Bilateral Damage to the Human Amygdala. *Nature, 372*, 669-672.

Aggleton, J. P., Burton, M. J., & Passingham, R. E. (1980). Cortical and Sub-Cortical Afferents to the Amygdala of the Rhesus-Monkey (Macaca-Mulatta). *Brain Research, 190*, 347-368.

Allison, T., Puce, A., McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends in Cognitive Sciences*, *4*, 267-278.

Altschuler, E. L., Vankov, A., Wang, V., Ramachandran, V. S., Pineda, J. A (1997). Person see, person do: human cortical electrophysiological correlates of monkey see monkey do cell. *Society for Neuroscience* Abstracts, 719.17.

Altschuler, E. L., Vankov, A., Hubbard, E. M., Roberts, E., Ramachandran, V. S. Pineda, J. A (2000). Mu wave blocking by observation of movement and its possible use as a tool to study theory of others minds. Society for *Neuroscience* Abstracts, *26*, 67.23.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science, 17*, 841-844.

Apperly, I. A., Samson, D., & Humphreys, G. W. (2009). Studies of adults can inform accounts of theory of mind development. *Developmental Psychology, 45*, 190-201.

Astafiev, S. V., Stanley, C. M., Shulman, G. L., & Corbetta, M. (2004). Extrastriate body area in human occipital cortex responds to the performance of motor actions. *Nature Neuroscience, 7*, 542-548.

Atkinson, A., & Adolphs, R. (2005). Visual emotion perception: mechanisms and processes. *Emotion and consciousness* (pp. 150-182). New York: Guilford Press.

Barbas, H. (1988). Anatomic Organization of Basoventral and Mediodorsal Visual Recipient Prefrontal Regions in the Rhesus-Monkey. *Journal of Comparative Neurology, 276*, 313-342.

Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer (Ed.). Advances in social cognition (Vol. 10, pp. 1–62). NJ: Mahwah: Erlbaum.

Baron-Cohen, S. (2006). The hyper-systemizing, assortative mating theory of autism. *Progress in Neuro-Psychopharmacology & Biological Psychiatry, 30*, 865-872.

Baron-Cohen, S. (2005). The empathising system: A revision of the 1994 model of the Mindreading System. In: Origins of the Social Mind. (B. Ellis and D. Bjorklund, Eds.), Guilford: Guilford Publications Inc.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences, 6,* 248-254.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.

Baron-Cohen, S. (1991a). Precursors to a theory of mind: Understanding attention in others. In A. Whiten (Ed.), *Natural theories of mind*. Oxford: Basil Blackwell.

Baron-Cohen, S. (1991b). Do people with autism understand what causes emotion? *Child Development*, *62*, 385-395.

Baron-Cohen, S. (1989). Joint attention deficits in autism: towards a cognitive analysis. *Development and Psychopathology*, 1, 185-189.

Baron-Cohen, S. (1987). Autism and symbolic play. *British Journal of Developmental Psychology*, *5*, 139-148.

Baron-Cohen, S., Campbell, R., Karmiloff-Smith, A., Grant, J., & Walker, J. (1995). Are children with autism blind to the mentalistic significance of the eyes. *British Journal of Developmental Psychology, 13*, 379-398.

Baron-Cohen, S., Jolliffe, T., Martimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry*, *38*, 813-822.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1986). Mechanical, behavioural and Intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology*, *4*, 113-125.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the Autistic-Child Have a Theory of Mind. *Cognition, 21*, 37-46.

Baron-Cohen, S., Ring, H., Bullmore, E., Wheelwright, S., Ashwin, C., & Williams, S. (2000). The amygdala theory of autism. *Neuroscience and Behavioural Reviews*, *24*, 355-364.

Baron-Cohen, S., Spitz, A., & Cross, P. (1993). Do Children with Autism Recognize Surprise - a Research Note. *Cognition & Emotion, 7*, 507-516.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The 'Reading the Mind in the eyes' test revised version: A study with normal adults, and adults with Asperger Syndrome or High-Functioning autism. *Journal of Child Psychiatry and Psychiatry, 42*, 241-252.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*, 5-17.

Behrmann, M., Thomas, C., & Humphreys, K. (2006). Seeing it differently: visual processing in autism. *Trends in Cognitive Sciences, 10*(6), 258-264.

Berger, H. (1929). *Über das Elektroenkephalogramm des Menschen.* Archiv für Psychiatrie und Nervenkrankheiten, 87, 527-570.

Best, C. A., Minshew, N. J., & Strauss, M. S. (2010). Gender Discrimination of Eyes and Mouths by Individuals with Autism. *Autism Research*, *3*, 88-93.

Blakemore, S. J., Boyer, P., Pachot-Clouard, P., Meltzoff, A., Segebarth, C., & Decety J. (2003). The detection of contingency and animacy from simple animations in the human brain. *Cerebral Cortex*, *13*, 837-844.

Blakemore, S. J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, *2*, 561-567.

Blakemore, S. J., Winston, J., & Frith, U. (2004). Social cognitive neuroscience: where are we heading? *Trends in Cognitive Sciences, 8*, 216-222.

Bloom, P., & German, T. (2000). Two reasons for abandoning the false belief task as a test of theory of mind. *Cognition*, *77*, 25-31.

Boria, S., Fabbri-Destro, M., Cattaneo, L., Sparaci, L., Sinigaglia, C., Santelli, E., et al. (2009). Intention Understanding in Autism. *PlosOne, 4*, e5596.

Botvinick, M., Jha, A. P., Bylsma, L. M., Fabian, S. A., Solomon, P. E., & Prkachin, K. M. (2005). Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. *NeuroImage, 25*, 312-319.

Boucher, J., & Lewis, V. (1992). Unfamiliar Face Recognition in Relatively Able Autistic-Children. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 33*, 843-859.

Boucher, J., Lewis, V., & Collis, G. (1998). Familiar face and voice matching and recognition in children with autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 39*, 171-181.

Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating action understanding: Inferential processes versus action simulation. *Current Biology, 17*, 2117-2121.

Brothers, L. 1990. The social brain: a project for integrating primate behavior and neurophysiology in a new domain. *Concepts Neuroscience*. *1*, 27-51.

Bruce, V. (1986). Recognising familiar faces. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. W. Young (Eds.), *Aspects of face processing* (pp. 107-117). Dordrecht: Martinus Nijhoff.

Bruce, V., & Young, A. (1986). Understanding Face Recognition. *British Journal of Psychology, 77*, 305-327.

Bruyer, R., Laterre, C., Seron, X., Feyereisen, P., Strypstein, E., Pierrard, E., et al. (1983). A Case of Prosopagnosia with Some Preserved Covert Remembrance of Familiar Faces. *Brain and Cognition, 2*, 257-284.

Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H. J., et al. (2004). Neural circuits underlying imitation learning of hand actions: An event-related fMRI study. *Neuron, 42*, 323-334.

Bushnell, I. W. R., Sai, F., & Mullin, J. T. (1989). Neonatal Recognition of the Mothers Face. *British Journal of Developmental Psychology, 7*, 3-15.

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology-Human Perception and Performance, 26*, 527-551.

Calder, A. J., Young, A. W., Rowland, D., Perrett, D. I., Hodges, J. R., & Etcoff, N. L. (1996). Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. *Cognitive Neuropsychology, 13*, 699-745.

Calvo, M. G., & Marrero, H. (2009). Visual search of emotional faces: The role of affective content and featural distinctiveness. *Cognition and Emotion*, 23, 782-806.

Calvo, M. G., & Nummenmaa, L. (2008). Detection of emotional faces: Salient physical features guide effective visual search. *Journal of Experimental Psychology-General*, *137*, 471-494.

Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex, 15*, 1243-1249.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development*, *21*, 315-330.

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences USA*, *100*, 5497-5502.

Cashdan, E. (2004) Smiles, speech, and body posture: How women and men display sociometric status and power. *Journal of Nonverbal Behavior,* 22, 209-28.

Castelli, F. (2005). Understanding emotions from standardized facial expressions in autism and normal development. *Autism*, *9*, 428-449.

Castelli, F., Frith, C., Happe, F., & Frith, U. (2002) Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain*, *125*, 1839-1849.

Castelli, F., Happe, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage*, *12*, 314-325.

Celani, G., Battacchi, M. W., & Arcidiacono, L. (1999). The understanding of the emotional meaning of facial expressions in people with autism. *Journal of Autism and Developmental Disorders*, *29*, 57-66.

Churchland, P. M. (1991). Folk psychology and the explanation of human behavior. In: J. D. Greenwood (Ed.). *The future of folk psychology*. Cambridge: Cambridge University Press. 51-69.

Clark, A. (1997). *Being There: Putting Brain, Body and World Together Again*. Cambridge MA: The MIT Press.

Clark, A., & Chalmers, D. (1998). The extended mind (Active externalism). *Analysis, 58*, 7-19.

Corden, B., Chivers, R., & Skuse, D. (2008). Avoidance of emotionally arousing stimuli predicts social-perceptual impairment in Asperger's syndrome. *Neuropsychologia*, *46*, 137-147.

Cross, E. S., Hamilton, A. F. D. C., & Grafton, S. T. (2006). Building a motor simulation de novo: Observation of dance by dancers. *NeuroImage, 31*, 1257-1267.

Cross, E. S., Kraemer, D. J. M., Hamilton, A. F. D., Kelley, W. M., & Grafton, S. T. (2009). Sensitivity of the Action Observation Network to Physical and Observational Learning. *Cerebral Cortex, 19*, 315-326.

Csibra, G. (2007). Action mirroring and action interpretation: An alternative account. In: P. Haggard, Y. Rossettu, & M. Kawato (Eds.), *Sensorimotor Foundations of Higher Cognition. Attention and Performance XXII* (pp.435-459). Oxford: Oxford University Press.

Csibra, G., Biro, S., Koos, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science, 27*, 111-133.

Csibra, G., & Gergely, G. (2007). 'Obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica, 124*, 60-78.

Csibra, G., & Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science*, *1*, 255-259.

Curcio, F. (1978). Sensorimotor functioning and communication in mute autistic children. *Journal of Autism and Developmental Disorders*, *8*, 281–292.

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., & Iacoboni, M. (2006). Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience*, *9*, 28-30.

Darwin, C. (1872/1998). The expression of the emotions in man and animals (3rd ed.). London: Harper Collins.

Davidson, R. J., & Hughdahl, K. (Eds.) (1995). *Brain asymmetry*. Cambridge, MA: MIT Press.

Davies, S., Bishop, D., Manstead, A. S. R., & Tantam, D. (1994). Face perception in children with autism and Asperger's syndrome. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *35*, 1033-1057.

Decety, J., & Chaminade, T. (2003). When the self represents the other: A new cognitive neuroscience view on psychological identification. *Consciousness and Cognition, 12*, 577-596.

Decety, J., & Grezes, J. (2006). The power of simulation: Imagining one's own and other's behavior. *Brain Research, 1079*, 4-14.

Decety, J., & Grezes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Sciences, 3*, 172-178.

Decety, J., Grezes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., et al. (1997). Brain activity during observation of actions - Influence of action content and subject's strategy. *Brain, 120*, 1763-1777.

De Lange, F.P., Spronk, M., Willems, R.M., Toni, I., Bekkering, H. (2008).Complementary systems for understanding action intentions. *Current Biology, 18*, 454-457.

Del Giudice, M., & Colle, L. (2007). Differences between children and adults in the recognition of enjoyment smiles. *Developmental Psychology*. 43, 796-803.

Dennett, D. C. (1987). *The Intentional Stance*. The MIT Press: Cambridge, MA.

Dimberg, U. (1982). Facial Reactions to Facial Expressions. *Psychophysiology, 19*, 643-647.

Dimberg U., Thunberg M. (1998), Rapid facial reactions to emotional facial expressions. *Scandinavian Journal of Psychology*, 39, 39-45.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science*, *11*, 86-89.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding Motor Events - a Neurophysiological Study. *Experimental Brain Research, 91*, 176-180.

Dittrich, W. H. (1993). Action Categories and the Perception of Biological Motion. *Perception*, *22*, 15-22.

Dittrich, W. H., Troscianko, T., Lea, S. E. G., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception*, *25*, 727-738.

Downing, P. E., Bray, D., Rogers, J., & Childs C. (2004). Bodies capture attention when nothing is expected. *Cognition*, *93*, B27-B28.

Downing, P. E., Chan, A. W. Y., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain specificity in visual cortex. *Cerebral Cortex*, *16*, 1453-1461.

Downing, P. E., Jiang, Y. H., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science, 293*, 2470-2473.

Downing, P., Liu, J., & Kanwisher, N. (2001). Testing cognitive models of visual attention with fMRI and MEG. *Neuropsychologia, 39*, 1329-1342.

Downing, P. E., & Peelen, M. V. (2011). The role of occipitotemporal body-selective regions in person perception. *Cognitive Neuroscience, 2*, 186-203.

Downing, P. E., Peelen, M. V., Wiggett, A. J., & Tew, B. D. (2006). The role of the extrastriate body area in action perception. *Social Neuroscience*, *1*, 52-62.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition, 6*, 509-540.

Eibl-Eibesfeldt, I. (1972) Similarities and differences between cultures in expressive movements. *Non-verbal communication*. Cambridge University Press.

Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia, 45*, 15-31.

Ekman, P. (2004). Emotions revealed: understanding faces and feelings. London: Phoenix.

Ekman, P. (2003). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. New York: Times Books/Henry Holt.

Ekman, P. (2001). *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. New York: W. W. Norton.

Ekman, P. (1994). Strong Evidence for Universals in Facial Expressions - a Reply to Russells Mistaken Critique. *Psychological Bulletin, 115*, 268-287.

Ekman, P. (1993). Facial Expression and Emotion. *American Psychologist, 48*, 384-392.

Ekman, P. (1992). Are There Basic Emotions. *Psychological Review, 99*, 550-553.

Ekman, P., & Friesen W. V. (1982). Felt, false and miserable smiles. *Journal of Nonverbal Behavior*, *6*, 238-252.

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System*, Consulting Psychologists Press, Palo Alto, CA.

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Consulting Psychologists Press. Palo Alto, CA.

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face.* Englewood Cliffs, NJ: Prentice-Hall.

Ekman, P., & Friesen, W. V. (1971). Constants Accross Cultures In The Face And Emotion. *Journal of Personality and Social Psychology*, *17*, 124-129.

Ekman, P., Friesen, W. V., & Ancoli, S. (1980). Facial signs of emotional experience. *Journal of Personality and Social Psychology*, *6*, 1125-1134.

Ekman, P., Hager, J. C., & Friesen, W. V. (1981). The symmetry of emotional and deliberate facial actions. *Psychophysiology*, *18*, 101-106.

Ekman, P., & Oster, H. (1979). Facial Expressions of Emotion. *Annual Review of Psychology, 30*, 527-554.

Ekman, P., & O'Sullivan, M. (1988). The role of context in interpreting facial expressions – Comment. *Journal of Experimental Psychology-General*, *117*, 86-88.

Ellamil, M., Susskind, J. M., & Anderson, A. K. (2008). Examinations of identity invariance in facial expressions adaptation. *Cognitive, Affective, & Behavioral Neuroscience*, *8*, 273-281.

Enns, J. T., Lleras, A., & Moore, M. C. (2010). Object updating: A force for perceptual continuity and scene stability in human vision. In R. Nijhawan (2010). *Problems of space and time in perception and action* (pp.503-520). Cambridge University Press.

Fadiga L., Fogassi L., Pavesi G., Rizzolatti G. *(*1995). Motor facilitation during action observation: a magnetic stimulation study. *Journal of Neurophysiology*, 73, 2608-2611.

Ferrari, P. F., Gallese, V., Rizzolatti, G., & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience, 17*, 1703-1714.

Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

Fodor, J. A. (1981). The Mind-Body Problem. *Scientific American, 244*, 114-123.

Fodor, J. A. (1968). The appeal to tacit knowledge in psychological explanation. *Journal of Philosophy,* 65, 627-640.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: From action organization to intention understanding. *Science, 308*, 662-667.

Fox, C. J., & Barton, J. J. (2007). What is adapted in face adaptation? The neural representations of expression in the human visual system. *Brain Research, 1127*, 80-89.

Frank, M. G. (2002). Smiles, lies, and emotion. In M. H. Abel (Ed.), *An empirical reflection on the smile* (pp.15-43). Lewiston, NY: Edwin Mellen Press.

Frank, M. G., & Ekman, P. (1993). Not all smiles are created equal: The differences between enjoyment and nonenjoyment smiles. *Humor, 6,* 9-26.

Frank, M. G., Ekman, P., & Friesen, W. V. (1993). Behavioral markers and recognizability of the smile of enjoyment. *Journal of Personality and Social Psychology*, *64*, 83-93.

Frey, S. H., & Gerry, V. E. (2006). Modulation of neural activity during observational learning of actions and their sequential orders. *Journal of Neuroscience, 26*, 13194-13201.

Freyd, J. J. (1987). Dynamic mental representations. *Psychological Review*, *94*, 427-438.

Freyd, J. J. (1983). The mental representation of movement when static stimuli are viewed. *Perception & Psychophysics, 33,* 575–581.

Freyd, J. J., & Finke, R. A. (1984). Representational Momentum. *Journal of Experimental Psychology: Learning Memory and Cognition*, *10*, 126-132.

Fridlund, A. (1994). *Human facial expression: An evolutionary view.* San Diego, CA US: Academic Press.

Friedman, O., & Leslie, A. M. (2004). A developmental shift in processes underlying successful belief-desire reasoning. *Cognitive Science, 28*, 963-977.

Frith, C. D., & Frith, U. (2008). Implicit and Explicit Processes in Social Cognition. *Neuron*, *60*, 503-510.

Frith, U. (2003). *Autism: Explaining the enigma*. Oxford: Blackwell.

Frith, U. (2001). Mind blindness and the brain in autism. *Neuron, 32*, 969-979.

Frith, U. (1996). Cognitive explanations of autism. *Acta Paediatrica, 85*, 63-68.

Frith, U. (1995). *Autism: Explaining the Enigma*. Oxford, Blackwell.

Frith, U. (1991). Asperger and his syndrome. In U. Frith (Ed.), *Autism and Asperger syndrome* (pp. 1-36). Cambridge: Cambridge University Press.

Frith, U., Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society London, B: Biological Science*, *358*, 459-473.

Frith, U., Morton, J., & Leslie, A. M. (1991). The Cognitive Basis of a Biological Disorder - Autism. *Trends in Neurosciences, 14*, 433-438.

Gallagher*,* S. *(*2007*).* Simulation trouble. *Social Neuroscience, 2*, 353-65.

Gallagher, S. (2005). *How the Body Shapes the Mind*. Oxford: Oxford University Press.

Gallagher, S., & Zahavi, D. (2008). *The Phenomenological Mind. An Introduction to Philosophy of Mind and Cognitive Science*. New York and London: Routledge.

Gallese, V. (2007). Before and below 'theory of mind': Embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society of London: Series B*, *362*, 659-669.

Gallese, V. (2006a). Mirror neurons and intentional attunement: Commentary on olds. *Journal of the American Psychoanalytic Association, 54*, 47-57.

Gallese, V. (2006b). Intentional attunement: a neurophysiological perspective on social cognition and its disruption in autism. *Brain Research*, *1079*, 15-24.

Gallese, V. (2001). The 'shared manifold' hypothesis - From mirror neurons to empathy. *Journal of Consciousness Studies, 8*, 33-50.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (2002). Action representation and the inferior parietal lobule. *Common Mechanisms in Perception and Action, 19*, 334-355.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*, 593-609.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, *2*, 493-501.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Science*, *8*, 396-403.

Gallese, V., & Lakoff, G. (2005). The Brain's Concepts: The Role of the Sensory-Motor System in Reason and Language. *Cognitive Neuropsychology*, *22*, 455-479.

Gallese, V., & Sinigaglia, C. (2011). What is so special about embodied simulation? *Trends in Cognitive Sciences, 15*, 512-519.

Gamond, L., George N., Hugueville L., Adam C., & Tallon-Baudry C. (2011). Early influence of prior experience on face perception, *NeuroImage*, *54*, 1415-26.

Gepner, B., Deruelle, C., & Grynfeltt, S. (2001). Motion and emotion: A novel approach to the study of face processing by young autistic children. *Journal of Autism and Developmental Disorders, 31*, 37-45.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: the naive theory of rational action. *Trends in Cognitive Sciences*, *7*, 287-292.

Gergely, G., Nadasdy, Z., Csibra, G., & Biro, S. (1995). Taking the International Stance at 12 Months of Age. *Cognition, 56*, 165-193.

Gibson, J. J. (1979). The Ecological Approach to Visual Perception. Boston: Houghton Mifflin.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience, 4*, 179-192.

Goldman, A. I. (2009). Mirroring, Simulating and Mindreading. *Mind and Language, 24*, 235-252.

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology and neuroscience of mindreading*. Oxford: Oxford University Press.

Goldman, A. I. (1989). Interpretation psychologized. *Mind and Language, 4*, 161-185.

Goldman, A. I., & Sripada, C. S. (2005). Simulationist models of face-based emotion recognition. *Cognition, 94*, 193-213.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*, 20-25.

Gopnik, A., & Meltzoff, A. (1997). *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.

Gopnik, A., & Wellman, H. (1994). The Theory Theory. In L. Hirschfield and S. Gelman (eds). *Mapping the Mind Domain Specificity in Cognition and Culture*. Cambridge University Press, 257-293.

Gopnik, A., & Wellman, H. (1992). "Why the Child's Theory of Mind Really is a Theory". *Mind and Language*, *7*, 145-172.

Gordon, R. M. (2004). Folk psychology as mental simulation. In N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*.

Gordon, R. M. (1995). Simulation Without Introspection or Inference From Me to You. In: M. Davies & T. Stone (Eds.). *Mental Simulation*. Oxford: Blackwell. 53-67.

Gosselin, P., Kirouac, G., & Dore, F. Y. (1997). Components and recognition of facial expressionsin the communication of emotion by actors. *Journal of Personality and Social Psychology*, *68*, 83-96.

Gosselin, P., Perron, M., Legault, M., & Campanella, P. (2002). Children’s and adults’ knowledge of the distinction between enjoyment and nonenjoyment smiles. *Journal of Nonverbal Behavior*, *26*, 83-108.

Grafton, S. T., Arbib, M. A., Fadiga, L., & Rizzolatti, G. (1996). Localization of grasp representations in humans by positron emission tomography: 2. Observation compared with imagination. *Experimental Brain Research, 112*, 103-111.

Grèzes, J., Armony, J. L., Rowe, J., & Passingham, R. E. (2003). Activations related to mirror and canonical neurones in the human brain: An fMRI study. *NeuroImage*, *18*, 928-937.

Grezes, J., Frith, C., & Passingham, R. E. (2004). Brain mechanisms for inferring deceit in the actions of others. *Journal of Neuroscience*, *24*, 5500-5505.

Grèzes, G., Wicker, S., Berthoz, B., & de Gelder, B. (2009). A failure to grasp the affective meaning of actions in autism spectrum disorder subjects. *Neuropsychologia, 47*, 1816-1825.

Gross, D., & Harris, P. L. (1988). False beliefs about emotion: Children's understanding of misleading emotional displays. *International Journal of Behavioral Development*, *11*, 475-488.

Grossman, J. B., Klin, A., Carter, A. S., & Volkmar, F. R. (2000). Verbal bias in recognition of facial emotions in children with Asperger syndrome. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*, 369-379.

Hamilton, A. F. D., 2009. Goals, intentions and mental states: challenges for theories of autism. *Journal of Child Psychology and Psychiatry, 50*, 881–892.

Hamilton, A. F. D., Brindley, R. M., & Frith, U. (2007). Imitation and action understanding in autistic spectrum disorders: how valid is the hypothesis of a deficit in the mirror neuron system? *Neuropsychologia*, *45*, 1859-1868.

Happé, F. (1999). Autism: Cognitive deficit or cognitive style? *Trends in Cognitive Sciences, 3*, 216-222.

Happé, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 36,* 5-25.

Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S., & Rizzolatti, G. (1998). Activation of human primary motor cortex during action observation: A neuromagnetic study. *Proceedings of the National Academy of Sciences of the United States of America, 95*, 15061-15065.

Hari, R., & Salmelin, R. (1997). Human cortical oscillations: A neuromagnetic view through the skull. *Trends in Neurosciences, 20*, 44-49.

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial Emotion Recognition in Autism Spectrum Disorders: A Review of Behavioral and Neuroimaging Studies. *Neuropsychology Review, 20*, 290-322.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry, 51*, 59-67.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*, 223-233.

Heal, J. (1986). “Replication and Functionalism”, in *Language, Mind, and Logic*, J. Butterfield (ed.), Cambridge: Cambridge University Press; reprinted in M. Davies and T. Stone (eds.), *Folk Psychology: the Theory of Mind Debate*, Oxford: Blackwell Publishers, 1995.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behaviour. *American Journal of Psychology*, 57, 243-259.

Herschbach, M. (2008). Folk psychological and phenomenological accounts of social perception. *Philosophical Explorations*, 11, 223-235.

Hess, U., Beaupre’, M. G. Cheung, N. (2002). Who to whom and why – cultural differences and similarities in the function of smiles. In: Millicent Abel (Ed.) *An empirical reflection on the smile*, pp. 187-216. NY: The Edwin Mellen Press.

Hess, U., & Kleck, R. E. (1990). Differentiating emotion elicited and deliberate emotional facial expressions. *European Journal of Social Psychology*, *20*, 369-385.

Hill, M. L., & Craig, K. D. (2002). Detecting deception in pain expressions: the structure of genuine and deceptive facial displays. *Pain, 98*, 135-144.

Hobson, R. P. (1986a). The autistic child's appraisal of expressions of emotion. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *27*, 321-342.

Hobson, R. P. (1986b). The autistic child’s appraisal of expressions of emotion: A further study. *Journal of Child Psychology and Psychiatry*, *27*, 671-680.

Hoehl, S., Palumbo, L., Heinisch, C., & Striano, T. (2008). Infants' attention is biased by emotional expressions and eye gaze direction. *Neuroreport, 19*, 579-582.

Hoekstra, R. A., Bartels, M., Cath, D. C., & Boomsma, D. I. (2008). Factor structure, reliability and criterion validity of the autism-spectrum quotient (AQ): A study in dutch population and patient groups. *Journal of Autism and Developmental Disorders, 38*, 1555-1566.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience, 3*, 80-84.

Hsu, S. M., & Young, A. W. (2004). Adaptation effects in facial expression recognition. *Visual Cognition*, *11*, 871-899.

Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: A review of the findings. *Psychonomic Bulletin & Review*, *12*, 822-851.

Hudson, M., Burnett, H. G., & Jellema, T. (2011). Anticipation of behavioural intentions in Asperger’s syndrome. *Journal of Autism and Developmental Disorders, 29,* 407-418.

Hudson, M., & Jellema, T. (2011). Resolving ambiguous behavioural intentions by means of involuntary prioritisation of gaze processing. *Emotion*, *11*, 681-686.

Hudson, M., Liu, C. H., & Jellema, T. (2009). Anticipating intentional actions: The effect of eye gaze direction on the judgement of head rotation. *Cognition*, *112*, 423-434.

Humphreys, K., Minshew, N., Leonard, G. L., & Behrmann, M. (2007). A fine-grained analysis of facial expression processing in highfunctioning adults with autism. *Neuropsychologia*, *45*, 685-695.

Hurley, C. M., & Frank, M. G. (2011). Executing Facial Control During Deception Situations. *Journal of Nonverbal Behavior, 35*, 119-131.

Iacoboni, M. (2009). Imitation, Empathy, and Mirror Neurons. *Annual Review of Psychology, 60*, 653-670.

Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience*, *7*, 942-951.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *Plos Biology, 3*, 529-535.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science, 286*, 2526-2528.

Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin*, *115*, 288-299.

Jabbi, M., Swart, M., & Keysers, C. (2007). Empathy for positive and negative emotions in the gustatory cortex. *NeuroImage, 34*, 1744-1753.

Jacob, P. (2008). What do mirror neurons contribute to human social cognition? *Mind & Language, 23*, 190-223.

Jacob, P., & Jeannerod, M. (2005). The motor theory of social cognition: a critique. *Trends in Cognitive Sciences, 9*, 21-25.

Jastorff J, Clavagnier S, Gergely G, Orban GA. (2010). Neural mechanisms of understanding rational actions: middle temporal gyrus activation by contextual violation. *Cerebral Cortex*, *21*, 318-329.

Jeannerod, M. (1994). The Representing Brain - Neural Correlates of Motor Intention and Imagery. *Behavioral and Brain Sciences, 17*, 187-202.

Jellema, T., Baker, C. I., Wicker, B., & Perrett, D. I. (2000). Neural representation for the perception of the intentionality of actions. *Brain and Cognition, 44*, 280-302.

Jellema, T., Lorteije, J., van Rijn, S., van 't Wout, M., de Haan, E., van Engeland, H., & Kemner, C. (2009). Involuntary interpretation of social cues is compromised in autism spectrum disorders. *Autism Research, 2*, 192-204.

Jellema, T., Maassen, G., & Perrett, D. I. (2004). Single cell integration of animate form, motion and location in the superior temporal cortex of the macaque monkey. *Cerebral Cortex, 14*, 781-790.

Jellema, T., Pecchinenda, A., Palumbo, L., & Tan, E. G. (2011). Biases in the perception and affective valence of neutral facial expressions induced by the immediate perceptual history. *Visual Cognition*, *19*, 616-634.

Jellema, T., & Perrett, D. (2007). Neural pathways of social cognition. In R. I. M. Dunbar & L. Barrett (Eds.), *The Oxford Handbook of Evolutionary Psychology*. Oxford/New York: Oxford University Press.

Jellema, T., & Perrett, D. I. (2005). Neural basis for the perception of goal-directed actions. In A. Easton and N. J. Emery (Eds.): *The Cognitive Neuroscience of Social Behaviour* Psychology Press, New York. Studies in Cognition Series. pp 81-112.

Jellema, T., & Perrett, D. I. (2003). Perceptual history influences neural responses to face and body postures. *Journal of Cognitive Neuroscience*, *15*, 961-971.

Jellema, T., & Perrett, D. I. (2002a). Coding of visible and hidden actions. *Common Mechanisms in Perception and Action, 19*, 356-380.

Jellema, T., & Perrett, D.I. (2002b). Coding of visible and hidden actions. Attention and Performance XIX, 356-380 In Prinz, W., and Hommel, B. (Eds.) *Common Mechanisms in Perception and Action: Attention and Performance, Vol. XIX*. Oxford: Oxford University Press.

Jenkins, A. C., & Mitchell, J. P. (2010). Mentalizing under Uncertainty: Dissociated Neural Responses to Ambiguous and Unambiguous Mental State Inferences. *Cerebral Cortex, 20*, 404-410.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophisics*, *14*, 201-211.

Jones, B. C., DeBruine, L. M., Little, A. C., Conway, C. A., & Feinburg, D. R. (2006). Integrating gaze direction and expression in preferences for attractive faces. *Psychological Science*, *17*, 588-591.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience, 3*, 759-763.

Keltner, D. (1995). Signs of Appeasement - Evidence for the Distinct Displays of Embarrassment, Amusement, and Shame. *Journal of Personality and Social Psychology, 68*, 441-454.

Keysar, B., Lin, S. H., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*, 25-41.

Keysers, C., & Gazzola, V. (2007). Integrating simulation and theory of mind: from self to social cognition. *Trends in Cognitive Sciences, 11*, 194-196.

Keysers, C., & Perrett, D. I. (2004). Demystifying social cognition: a Hebbian perspective. *Trends in Cognitive Sciences, 8*, 501-507.

Keysers, C., Wicker, B., Gazzola, V., Anton, J. L., Fogassi, L., & Gallese, V. (2004). A touching sight: SII/PV activation during the observation and experience of touch. *Neuron, 42*, 335-346.

Kilner, J. M. (2009). Dissociable functional roles of the human action-observation network (Commentary on E. S. Cross et al.). *European Journal of Neuroscience, 30*, 1382-1382.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). The mirror-neuron system: a Baynesian perspective. *Neuroreport, 18*, 619-623.

Kleinman, J., Marciano, P. L., & Ault, R. L. (2001). Advanced theory of mind in high-functioning adults with autism. *Journal of Autism and Developmental Disorders, 31*, 29-36.

Kohler, C.G., Turner, T.T., Bilker, W.B., Brensinger, C., Siegel, S.J., Kanes, S.J., Gur, R.E., Gur, R.C., 2003. Facial emotion recognition in schizophrenia: intensity effects and error pattern. *American Journal of Psychiatry, 160*, 1168-1174.

Kohler, C. G., Turner, T., Stolar, N.M., Bilker,W. B., Brensinger, C. M., Gur, R. E., et al. (2004). Differences in facial expressions of four universal emotions. *Psychiatry Research*, *128*, 235-244.

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science, 297*, 846-848.

Kovacs, G., Cziraki, C., Vidnyanszky, Z., Schweinberger, S. R., & Greenlee, M. W. (2008). Position-specific and position-invariant face aftereffects reflect the adaptation of different cortical areas. *NeuroImage*, *43*, 156-164.

Kozak, M. N., Marsh, A. A., & Wegner, D. M. (2006). What do I think you're doing? Action identification and mind attribution. *Journal of Personality and Social Psychology*, *90*, 543-55.

Kozlowski, L. T., Cutting, J. E. (1977). Recognizing sex of a walker from a dynamic point-light display. *Perception and Psychophisics, 21*, 575-580.

Kraut, R . E .. & Iohnston, R.E. (1979). Social and emotional messages of smiling; An ethological approach. *Journal of Personality and Social Psychology*, *37*, 1539-1553.

Krumhuber, E., Manstead, A., & Kappas, A. (2007). Temporal aspects of facial displays in person and expression perception: The effects of smile dynamics, head-tilt, gender. *Journal of Nonverbal Behavior, 31,* 39-56.

Kuchinke, L., Schneider, D., Kotz, S. A., & Jacobs, A. M. (2011). Spontaneous but not explicit processing of positive sentences impaired in Asperger's syndrome: Pupillometric evidence. *Neuropsychologia*, *49*, 331-338.

Kuzmanovic, B., Schilbach, L., Lehnhardt, F. G., Bente, G., & Vogeley, K. (2011). A matter of words: Impact of verbal and nonverbal information on impression formation in high-functioning autism. *Research in Autism Spectrum Disorders*, *5*, 604-613.

Lakoff, G., & Johnson, M. (1999). *Philosophy In The Flesh: the Embodied Mind and its Challenge to Western Thought*. Basic Books.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention and the startle reflex. *Psychological Review*, 97, 377-395.

Larsen, J. T., Norris, C. J., Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, *40*, 776-785.

Law Smith, M. J., Montagne, B., Perrett, M. G., & Gallagher, L. (2010). Detecting subtle facial emotion recognition deficits in high-functioning autism using dynamic stimuli of varying intensities. *Neuropsychologia*, *48*, 2777-2781.

Leekam, S., Baron-Cohen, S., Perrett, D., Milders, M., & Brown, S. (1997). Eye-direction detection: A dissociation between geometric and joint attention skills in autism. *British Journal of Developmental Psychology*, *15*, 77-95.

Leekam, S. R., & Perner, J. (1991). Does the Autistic-Child Have Metarepresentational Deficit. *Cognition, 40*, 203-218.

Lennon, K. (1990). *Explaining Human Action*. London: Duckworth, Chicago: Open Court.

Lennon, K., Burwood, S., & Gilbert, P. (2003). *Philosophy of Mind*. London: UCL Press, 1998. Reprinted 2003 by Routledge.

Lennon, K., & Gilbert, P. (2005). *The World, the Flesh and the Subject: Continental Themes in Philosophy of Mind and Body*, Edinburgh: Edinburgh University Press.

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*, *4*, 89-94.

Leslie, A. M. (1994a). ToMM, ToBy, and Agency: Core architecture and domain specificity. In L.A. Hirschfeld, & S. Gelman, (Eds), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119-148). New York: Cambridge University Press.

Leslie, A. M. (1994b). Pretending and believing - Issues in the theory of ToMM. *Cognition*, *50*, 211-238.

Leslie, A. M, & Roth, D. (1993). What can autism teach us about metarepresentation? In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen

(Eds.), *Understanding other minds: perspectives from autism*, Oxford: Oxford Medical Publications.

Leslie, K. R., Johnson-Frey, S. H., & Grafton, S. T. (2004). Functional imaging of face and hand imitation: towards a motor theory of empathy. *NeuroImage, 21*, 601-607.

Lewis, V., & Boucher, J. (1988). Spontaneous, instructed and elicited play in relatively able autistic children. *British Journal of Developmental Psychology*, *6*, 325-339.

Lewis, C., Freeman, N. H., Hagestadt, C., & Douglas, H. (1994). Narrative Access and Production in Preschoolers False Belief Reasoning. *Cognitive Development, 9*, 397-424.

Lieberman, M. D. (2007). Social cognitive neuroscience: A review of core processes *Annual Review of Psychology*, *58*, 259-289.

Lindner, J. L., & Rosen, L. A. (2006). Decoding of emotion through facial expression, prosody and verbal content in children and adolescents with Asperger's syndrome. *Journal of Autism and Developmental Disorders*, *36*, 769-777.

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (1999). *Autism Diagnostic Observation Schedule*. Los Angeles, CA.: Western Psychological Services.

Lou, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science*, *16*, 601-608.

Loveland, K. A., & Landry, S. H. (1986). Joint Attention and Language in Autism and Developmental Language Delay. *Journal of Autism and Developmental Disorders, 16*, 335-349.

Loveland, K. A., TunaliKotoski, B., Chen, Y. R., Ortegon, J., Pearson, D. A., Brelsford, K. A., et al. (1997). Emotion recognition in autism: verbal and nonverbal information. *Development and Psychopathology*, *9*, 579–593.

Macintosh, K. E., & Dissanayake, C. (2004). Annotation: The similarities and differences between autistic disorder and Asperger's disorder: a review of the empirical evidence. *Journal of Child Psychology and Psychiatry, 45*, 421-434.

Maringer, M., Krumhuber, E. G., Fischer, A. H., & Niedenthal, P. M. (2011). Beyond Smile Dynamics: Mimicry and Beliefs in Judgments of Smiles. *Emotion, 11*, 181-187.

Marsh, L. E., &, Hamilton, A. F. D. (2011). Dissociation of mirroring and mentalizing systems in autism. *NeuroImage*, *56*, 1511-1519.

McIntosh, D. N., Reichmann-Decker, A., Winkielman, P., & Wilbarger, J. L. (2006). When the social mirror breaks: Deficits in automatic, but not voluntary mimicry of emotional facial expressions in autism. *Developmental Science, 9*, 295-302.

McPartland, J. & Klin, A. (2006). Asperger Syndrome. *Adolescent Medicine Clinics of North America, 17,* 771-778.

Meltzoff, A. N. (1995). Understanding the intentions of others – re-enactment of intended acts by 18-month-old children, *Developmental Psychology*, *31*, 838-850.

Merleau-Ponty, M. (1964). *The primacy of perception and other essays on phenomenological psychology.* Evanston, IL: Nortwestern University Press.

Merleau-Ponty, M. (1962). *Phenomenology of Perception*. Translated by Colin Smith. Atlantic Highlands, New Jersey: The Humanities Press. Original edition, 1945.

Mermillod, M., Maringer, M., & Hess, U., & Niedenthal, P. M. (2010).The Simulation of Smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavioral and Brain Sciences*, *33*, 417-480.

Michotte, A. (1963). *The perception of causality* (trans. T. R. Miles & E. Miles). New York: Basic Books.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the Structure of Behavior*. New York: Holt, Rinehart & Winston.

Miller, J. N., & Ozonoff, S. (2000). The external validity of Asperger disorder: Lack of evidence from the domain of neuropsychology. *Journal of Abnormal Psychology, 109*, 227-238.

Minio-Paluello, I., Baron-Cohen, S., Avenanti, A., Walsh, V., & Aglioti, S. M. (2008). *Biological Psychiatry*, *65*, 55-62.

Mitchell, P., & Lacohee, H. (1991). Childrens Early Understanding of False Belief. *Cognition, 39*, 107-127.

Moore, C., Jarrold, C., Russell, J., Lumb, A., Sapp, F., & MacCallum, F. (1995). Conflicting desire and the child's theory of mind. *Cognitive Development, 10*, 467-482.

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., et al. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature, 383*, 812-815.

Morton, J., & Johnson, M. H. (1991). Conspec and Conlern - a 2-Process Theory of Infant Face Recognition. *Psychological Review, 98*, 164-181.

Mottron, L., Dawson, M., Soulieres, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders, 36*, 27-43.

Mühlberger, A., Wieser, M. J., Gerdes, A. B. M., Frey,nM. C. M., Peter Weyers, P., & Paul Pauli, P. (2011). Stop looking angry and smile, please: start and stop of the very same facial expression differentially activate threat- and reward-related brain networks. *Social, Cognitive and Affective Neuroscience, 6*, 321-329.

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-Neuron Responses in Humans during Execution and Observation of Actions. *Current Biology, 20*, 750-756.

Mundy, P., Sigman, M., Ungerer, J., & Sherman, T. (1986). Defining the Social Deficits of Autism - the Contribution of Nonverbal-Communication Measures. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 27*, 657-669.

Nakamura, K., Kawashima, R., Ito, K., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., Nagumo, S., Kubota, K., Fukuda, H., & Kijima, S. (1999). Activation of the right inferior frontal cortex during assessment of facial emotion. *Journal of Neurophysiology*, *82*, 1610-1614.

Naumann, L. P., Vazire, S., Rentfrow, P. J., & Gosling, S. D. (2009). Personality Judgments Based on Physical Appearance. *Personality and Social Psychology Bulletin, 35*, 1661-1671.

Neta, M., Whalen, P. J. (2010). The Primacy of Negative Interpretations When Resolving the Valence of Ambiguous Facial Expressions. *Psychological Science*, *21*, 901-907.

Newell, L. C., Best, C. A., Gastgeb, H., Rump, K. M., & Strauss, M. S. (2011). *The development of categorization and facial knowledge: Implications for the study of autism*. In: L. M. Oakes, C. H. Cashon, M. Casasola, D. H. Rakison editors. Early Perceptual and Cognitive Development. New York: Oxford Press.

Niedenthal, P. M. (2007). Embodying emotion. *Science, 316*, 1002-1005.

Niedenthal, P. M., Mermillod, Maringer, M., & Hess, U. (2010). The Simulation of Smiles (SIMS) Model: Embodied simulation and the meaning of facial expression, *Brain and Behavioural Sciences, 33*, 417-480.

Nijboer, T. C. W., & Jellema, T. (2012). Unequal impairments in the recognition of positive and negative emotions after right hemisphere lesions: A left hemisphere bias for happy faces. *Journal of Neuropsychology*, *6*, 79-93.

Nishitani, N., Avikainen, S., & Hari, R. (2004). Abnormal imitation-related cortical activation sequences in Asperger's syndrome. *Annals of Neurology*, *55*, 558-562.

Oberman, L. M., Hubbard, E. M., McCleery, J. P., Altschuler, E. L., Ramachandran, V. S., & Pineda, J. A. (2005). EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Cognitive Brain Research, 24*, 190-198.

Oberman, L., Winkielman, P., & Ramachandran, V. (2007). Face to face: Blocking facial mimicry can selectively impair recognition of emotional expressions. *Social Neuroscience, 2*, 167-178.

O'Connor, K., Hamm, J. P., & Kirk, I. J. (2005). The neurophysiologic correlates of face processing in adults and children with Asperger's syndrome. *Brain and Cognition*, *59*, 82-95.

Ogai, M., Matsumoto, H., Suzuki, K., Ozawa, F., Fukuda, R., Uchiyama, I., et al. (2003). fMRI study of recognition of facial epressions in high-functioning autistic patients. *Neuroreport*, *14*, 559-563.

Ohman, A. (1993). Fear and anxiety as emotional phenomenon: Clinical phenomenology, Evolutionary perspectives, and information-processing mechanisms. In M. Lewis & J. M. Haviland (Eds.). *Handbook of emotions* (pp. 511-536). New York: Guildford Press.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*, 255-258.

Onishi, K. H., Baillargeon, R., & Leslie, A. M. (2007). 15-month-old infants detect violations in pretend scenarios. *Acta Psychologica, 124*, 106-128.

Oosterhof, N. N., & Todorov, A. (2009). Shared Perceptual Basis of Emotional Expressions and Trustworthiness Impressions From Faces. *Emotion, 9*, 128-133.

Ortigue, S., Sinigaglia, C., Rizzolatti, G., Grafton, S. T. (2010). Understanding actions of others: the electrodynamics of the left and right hemispheres. A high-density EEG neuroimaging study. *PlosOne*, *5*, e12160.

Otta, E., Abrosio, F. F. E., & Hoshino, R. L. (1996). Reading a smiling face: Messages conveyed by various forms of smiling. *Perceptual and Motor Skills, 82*, 1111-1121.

Otta, E., Lira, B. B. P., Delevati, N. M., Cesar, O. P., & Pires, C. S. G. (1994). The Effect of Smiling and of Head Tilting on Person Perception. *Journal of Psychology, 128*, 323-331.

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1990). Are there emotion perception deficits in young autistic children? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *31*, 343-361.

Palermo, R., & Rhodes, G. (2007). Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia, 45,* 75-92.

Palmer, M. T., & Simmons, K. B. (1995). Communicating intentions through nonverbal behaviours: Conscious and nonconscious encoding of liking. *Human Communication Research*, *22*, 128-160.

Palumbo, L., & Jellema, T. (under review). Beyond face value: involuntary emotional anticipation shapes the perception of dynamic facial expressions.

Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience*, *8*, 636-648.

Pelphrey, K. A., Morris, J. P., McCarthy, G., & Labar, K. S. (2007). Perception of dynamic changes in facial affect and identity inautism. *Social Cognitive and Affective Neuroscience*, *2*, 140-149.

Pelphrey, K. A., Morris, J. P., Michelich, C. R., Allison, T., & McCarthy, G. (2005). Functional anatomy of biological motion perception in posterior temporal cortex: An fMRI study of eye, mouth and hand movements, *Cerebral Cortex*, *15*, 1866-1876.

Pelphrey, K. A., Singerman, J. D., Allison, T., & McCarthy, G. (2003). Brain activation evoked by perception of gaze shifts: the influence of context, *Neuropsychologia*, *41*, 156-170.

Perner, J., Frith, U., Leslie, A. M., & Leekam, S. R. (1989). Exploration of the Autistic Childs Theory of Mind - Knowledge, Belief, and Communication. *Child Development, 60*, 689-700.

Perrett, D. I. (1999). A cellular basis for reading minds from faces and actions. *Design of Animal Communication*, 159-185.

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and Functions of Cells Responsive to Faces in the Temporal Cortex. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 335*, 23-30.

Perrett, D. I., May, K. A., & Yoshikawa, S. (1994). Facial shape and judgements of female actractiveness. *Nature*, *368*, 239-242.

Perrett, D. I., Xiao, D., Barraclough, N. E., Keysers, C., & Oram, M. W. (2009). Seeing the future: natural image sequences produce ‘anticipatory’ neuronal activity and bias perceptual report. *Quarterly Journal of Experimental Psychology, 62*, 2081-2104.

Perrett, D. I., Xiao, D., Jellema, T., Barraclough, N., & Oram, M. W. (2006). Social perception from static and dynamic visual information. *Perception, 35*, 120-120.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion*: *Central and peripheral routes to attitude change*. New York: Springer-Verlag.

Pitcher, D., Garrido, L., Walsh, V., & Duchaine, B. C. (2008). Transcranial magnetic stimulation disrupts the perception and embodiment of facial expressions. *Journal of Neuroscience, 28*, 8929-8933.

Pons, F., Harris, P. L., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology*, *1*, 127-152.

Porter, S., & ten Brinke, L. (2008). Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. *Psychological Science, 19*, 508-514.

Premack, D. (1990). The infants theory of self-propelled objects. *Cognition*, *36*, 1-16.

Press, C., Richardson, D. C., & Bird, G. (2010). Intact imitation of emotional facial actions in autism spectrum conditions, *Neuropsychologia*, *48*, 3291-3297.

Prior, M., Dahlstrom, B., & Squires, T. L. (1990). Autistic children's knowledge of thinking and feeling states in other people. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *31*, 587-601.

Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *Journal of Neuroscience, 18*, 2188-2199.

Pylyshyn, Z. W. (1987). Whats in a Mind. *Synthese, 70*, 97-122.

Read, S. J., Druian, P. R., & Miller L. C. (1989). The role of causal sequence in the meaning of actions. *The British Journal of Social Psychology*, *28*, 341-351.

Reed, T., & Peterson, C. C. (1990). A comparative study of autistic subjects' performance at two levels of visual and cognitive perspective-taking. *Journal of Autism and Developmental Disorders*, *20*, 555-567.

Reis, H. T., Wilson, I. M., Monestere, C., Bernstein, S., Clark, K., Seidl, et al. (1990). What is smiling is beautiful and good. *European Journal of Social Psychology*, *20*, 259-267.

Rhodes, G., Jeffery, L., Watson, T. L., Clifford, C. W. G., & Nakayama, K. (2003). Fitting the mind to the world: Face adaptation and attractiveness aftereffects. *Psychological Science, 14*, 558-566.

Ristic, J., Mottron, L., Friesen, C. K., Iarocci, G., Burack, J. A., & Kingstone, A. (2005). Eyes are special but not for everyone: The case of autism. *Cognitive Brain Research, 24*, 715-718.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169-192.

Rizzolatti, G., & Fabbri-Destro, M. (2008). The mirror system and its role in social cognition. *Current Opinion in Neurobiology, 18*, 179-184.

Rizzolatti, G., Fabbri-Destro, M., & Cattaneo, L. (2009). Mirror neurons: from discovery to autism. *Experimental Brain Research*. *200*, 223-237.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*, 131-141.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience, 2*, 661-670.

Rochat, M. J., Serra, E., Fadiga, L., & Gallese, V. (2008). The evolution of social cognition: Goal familiarity shapes monkeys' action understanding. *Current Biology, 18*, 227-232.

Rozzi, S., Calzavara, R., Belmalih, A., Borra, E., Gregoriou, G. G., Matelli, M., & Luppino G. (2006). Cortical connections of the inferior parietal cortical convexity of the macaque monkey. *Cerebral Cortex*, *16*, 1389-1417.

Ruffman, T., & Perner, J. (2005). Do infants really understand false belief? *Trends in Cognitive Sciences, 9*, 462-463.

Rumsey, J. M., & Hamburger, S. D. (1988). Neuropsychological Findings in High-Functioning Men with Infantile-Autism, Residual State. *Journal of Clinical and Experimental Neuropsychology, 10*, 201-221.

Russell, J. (1997). How executive disorders can bring about an inadequate 'theory of mind' In J. Russell (Ed). *Autism as an executive disorder*. (pp. 256-304). New York, NY, US: Oxford University Press.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1161-1178.

Russell, J. A., & Bullock, M. (1986). On the dimensions preschoolers use to interpret facial expressions of emotions. *Developmental Psychology*, *22*, 97-102.

Russell, J. A., & Bullock, M. (1985). Multidimensional scaling of emotional facial expressions: similarity from preschoolers to adults. *Journal of Personality and Social Psychology*, *48*, 1290-1298.

Russel J. A., Fehr B. (1987). Relativity in the perception of emotion in facial expressions. *Journal of experimental psychology, 116*, 223-237.

Russell, J. A., Fernández-Dols, J. M. (1997). *The Psychology of Facial Expression*. Eds. New York: Cambridge University.

Rutherford, M. D., Chattha, H. M., & Krysko, K. M. (2008). The use of aftereffects in the study of relationships among emotion categories. *Journal of Experimental Psychology-Human Perception and Performance*, *34*, 27-40.

Rutherford, M. D., Pennington, B. F., & Rogers, S. J. (2006). The perception of animacy in young children with autism. *Journal of Autism and Developmental Disorders, 36*, 983-992.

Rutter, M. (2005). Incidence of autism spectrum disorders: Changes over time and their meaning. *Acta Paediatrica*, *94*, 2-15.

Rutter, M. (1978). Diagnosis and Definition of Childhood Autism. *Journal of Autism and Childhood Schizophrenia, 8*, 139-161.

Samson, D., & Apperly, I. A. (2010). There is More to Mind Reading than having Theory of Mind Concepts: New Directions in Theory of Mind Research. *Infant and Child Development, 19*, 443-454.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Scott, S. E. B. (2010). Seeing It Their Way: Evidence for Rapid and Involuntary Computation of What Other People See. *Journal of Experimental Psychology-Human Perception and Performance, 36*, 1255-1266.

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. *Reviews in the Neurosciences, 14*, 303-316.

Sander, D., Grandjean, D., Kaiser, S., Wehrle, T., & Scherer, K. R. (2007). Interaction effects of perceived gaze direction and dynamic facial expression: Evidence for appraisal theories of emotion. *European Journal of Cognitive Psychology*, *19*, 470-480.

Sato, W., Fujimura, T., & Suzuki, N. (2008). Enhanced facial EMG activity in response to dynamic facial expressions. *International Journal of Psychophysiology, 70*, 70-74.

Sato, W., Kochiyama, T., Yoshikawa, S., Naito, E., & Matsumura, M. (2004). Enhanced neural activity in response to dynamic facial expressions of emotion: an fMRI study. *Cognitive Brain Research, 20*, 81-91.

Sato, W., & Yoshikawa, S. (2004). The dynamic aspects of emotional facial expressions. *Cognition and Emotion*, *18*, 701-710.

Satpute, A. B., & Lieberman, M. D. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Research*, *1079*, 86-97.

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, *16*, 235-239.

Saxe, R. (2005). Against simulation: the argument from error. *Trends in Cognitive Sciences, 9*, 174-179.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology, 55*, 87-124.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people - The role of the temporo-parietal junction in "theory of mind". *NeuroImage, 19*, 1835-1842.

Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia, 43*, 1391-1399.

Saxe, R., Xiao, D. K., Kovacs, G., Perrett, D. I., & Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia*, *42*, 1435-1446.

Scherer, K. R. (1992). Emotions are biologically and socially constituted: A response to Greenwood. *New Ideas in Psychology, 10,* 19-22.

Schienle, A., Stark, R., & Walter, B. (2002). The insula is not specifically involved in disgust processing: an fMRI study. *Neuroreport, 13*, 2023-2026.

Schmidt, K. L., Ambadar, Z., Cohn, J. F., & Reed, L. I. (2006). Movement differences between deliberate and spontaneous facial expressions: Zygomaticus Major action in smiling. *Journal of Nonverbal Behavior*, *30*, 37-52.

Schmidt, K. L., & Cohn, J. F. (2001). Human facial expressions as adaptations: Evolutionary questions in facial expression research. *Yearbook of Physical Anthropology, 44*, 3-24.

Scholl, B. J., & Leslie, A. M. (1999). Modularity, development and 'theory of mind'. *Mind & Language, 14*, 131-153.

Senju, A., Southgate, V., Miura, Y., Matsui, T., Hasegawa, T., Tojo, Y., et al. (2010). Absence of spontaneous action anticipation by false belief attribution in children with autism spectrum disorder. *Development and Psychopathology, 22*, 353-360.

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous Theory of Mind in Asperger syndrome. *Science, 325*, 883-885.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701-703.

Siegal, M., & Beattie, K. (1991). Where to Look 1st for Childrens Knowledge of False Beliefs. *Cognition, 38*, 1-12.

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 809-813.

Slater, A., & Quinn, P. C. (2001). Face recognition in the newborn infant. *Infant and Child Development, 10*, 21-24.

Soussignan, R. (2002). Duchenne smile, emotional experience, and autonomic reactivity: A test of the facial feedback hypothesis. *Emotion*, *2*, 52-74.

Southgate, V., Gergely, G., & Csibra, G. (2010). Does the mirror neuron system and its impairment explain human imitation and autism? In: Pineda, J.A. (Ed.), *The Role of Mirroring Processes in Social Cognition*. Humana Press.

Southgate, V., & Hamilton, A. F. D. (2008). Unbroken mirrors: challenging a theory of Autism. *Trends in Cognitive Science*, *30*, 1-5.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science, 18*, 587-592.

Spengler, S., Bird, G., & Brass, M. (2010) Hyper-imitation of actions is related to reduced understanding of others' minds in autism spectrum conditions. *Biological Psychiatry*, *68* , 1148-1155.

Spiridon, M., & Kanwisher, N. (2002). How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron, 35*, 1157-1165.

Stich, S., & Nichols, S. (1992). Folk psychology: Simulation or tacit theory? *Mind and Language*, *7*, 35-71.

Striano, T., & Stahl, D. (2005). Sensitivity to triadic attention in early infancy. *Developmental Science, 8*, 333-343.

Surakka, V., & Hietanen, J. K. (1998). Facial and emotional reactions to Duchenne and non-Duchenne smiles. *International Journal of Psychophysiology*, *29*, 23-33.

Suzuki, S., & Cavanagh, P. (1998). A shape-contrast effect for briefly presented stimuli. *Journal of Experimental psychology: Human perception and performance*, *24*, 1315-1341.

Swettenham, J. (1996). Can children with autism be taught to understand false belief using computers? *Journal of Child Psychology and Psychiatry and Allied Disciplines, 37*, 157-165.

Swettenham, J., Baron-Cohen, S., Gomez, J.-C., & Walsh, S. (1996). What's inside a person's head? Conceiving of the mind as a camera helps children with autism develop an alternative theory of mind. *Cognitive Neuropsychiatry*, *1*, 73-88.

Tanaka-Matsumi, J., Attivissimo, D., Nelson, S., & D'urso, T. (1995). Context effects on the judgment of basic emotions in the face. *Motivation and Emotion*, *19*, 139-155.

Tantam, D., Monaghan, L., Nicholson, H., & Stirling, J. (1989). Autistic children's ability to interpret faces: A research note. *Journal of Child Psychology and Psychiatry*, *30*, 623-630.

Tassinary, L. G., & Cacioppo, J. T. (1992). Unobservable facial actions and emotion. *Psychological Science*, *3*, 28-33.

Taylor, J. C., Wiggett, A. J., & Downing, P. E. (2007). Functional MRI analysis of body and body part representations in the extrastriate and fusiform body areas. *Journal of Neurophysiology, 98*, 1626-1633.

Teufel, C., Alexis, D. M., Todd, H., Lawrance-Owen A. J., Clayton, N. S., Davis, G. (2009). Social cognition modulates the sensory coding of observed gaze direction. *Current Biology, 19*, 1274-1277.

Teufel, C., Fletcher, P. C., & Davis, G. (2010). Seeing other minds: attributed mental states influence perception. *Trends in Cognitive Science*, *14*, 376-382.

Teunisse, J. P., & de Gelder, B. (2001). Impaired categorical perception of facial expressions in high-functioning adolescents with autism. *Child Neuropsychology*, *7*, 1-14.

Thayer, S. (1980). The effect of expression sequence and expressor identity on judgments of intensity of facial expression. *Journal of Nonverbal Behavior*, *5*, 71-79.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

Thornton, I. M., & Vuong, Q. C. (2004). Incidental processing of biological motion, *Current Biology*, *14*, 1084-1089.

Tomasello, M., & Haberl, K. (2003). Understanding attention: 12- and 18-month-olds know what's new for other persons. *Developmental Psychology*, *39*, 906–912.

Tranel, D., Damasio, A. R., & Damasio, H. (1988). Intact Recognition of Facial Expression, Gender, and Age in Patients with Impaired Recognition of Face Identity. *Neurology, 38*, 690-696.

Uddin, L. Q., Iacoboni, M., Lange, C., & Keenan, J. P. (2007). The self and social cognition: the role of cortical midline structures and mirror neurons. *Trends in Cognitive Sciences, 11*, 153-157.

Umilta, M. A., Escola, L., Intskirveli, I., Grammont, F., Rochat, M., Caruana, F., et al. (2008). When pliers become fingers in the monkey motor system. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 2209-2213.

Umilta, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., et al. (2001). I know what you are doing: A neurophysiological study. *Neuron, 31*, 155-165.

Ungerer, J., & Sigman, M. (1981). Symbolic play and language comprehension in autistic children. *Journal of the American Academy of Child Psychiatry*, *20*, 318-337.

Uono, S., Sato, W., & Toichi, M. (2010). Brief report: Representational momentum for dynamic facial expressions in pervasive developmental disorder. *Journal of Autism and Developmental Disorders, 40*, 371-377.

Urgesi, C., Candidi, M., Ionta, S., & Aglioti, S. M. (2007). Representation of body identity and body actions in extrastriate body area and ventral premotor cortex. *Nature Neuroscience, 10*, 30-31.

Valenza, E., Simion, F., Cassia, V. M., & Umilta, C. (1996). Face preference at birth. *Journal of Experimental Psychology-Human Perception and Performance, 22*, 892-903.

Van Elk, M., van Schie, H. T., Hunnius, S., Vesper, C., & Bekkering, H. (2008). You'll never crawl alone: Neurophysiological evidence for experience-dependent motor resonance in infancy. *NeuroImage, 43*, 808-814.

Van Overwalle, F. & Baetens, K. (2009). Understading others' actions and goals by mirror and mentalyzing systems: A meta-analysis. *NeuroImage*, *48*, 564-584.

Van Rijsbergen, N., Jannati, A., & Treves, A. (2008). Aftereffects in the Perception of Emotion Following Brief, Masked Adaptor Faces. *The Open Behavioural Science Journal*, *2*, 36-52.

Venter, A., Lord, C., & Schopler, E. (1992). A Follow-up-Study of High-Functioning Autistic-Children. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 33*, 489-507.

Verbeke, E., Peeters, W., Kerkhof, I., Bijttebier, P., Steyaert, J., & Wagemans, J. (2005). Lack of motivation to share intentions: Primary deficit in autism? *Behavioral and Brain Sciences, 28*, 718-719.

Vinson, N. G., & Reed, C. L. (2002). Sources of object-specific effects in representational momentum. *Visual Cognition, 9*, 41-65.

Volkmar, F. R., Lord, C., Bailey, A., Schultz, R. T, Klin, A. (2004). Autism and pervasive developmental disorders. *Journal of Child Psychology and Psychiatry*, *45*, 135-170.

Wainer, B., & Handel, S. J. (1985). A cognition emotion action sequence – anticipated emotional consequences of causal attributions and reported communication strategy. *Developmental Psychology*, *21*, 102-107.

Walk, R. D., & Homan, C. P. (1984). Emotion and dance in dynamic-light displays, *Bullettin of the Psychonomic Society, 22*, 437-440.

Wang, A. T., Dapretto, M., Hariri, A. R., Sigman, M., & Bookheimer, Y. (2004). Neural correlates of facial affect processing in children and adolescents with autism spectrum disorder. *Journal f the American Academy of Child and Adolescent Psychiatry*, *43*, 481-490.

Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, *428*, 557-561.

Webster, M. A., & MacLin, O. H. (1999). Figural aftereffects in the perception of faces. *Psychonomic Bulletin & Review, 6*, 647-653.

Weiner, B., & Handel, S. (1986). A cognition-emotion-action sequence: anticipated emotional consequences of causal attribution and reported communication strategy. *Developmental psychology*, *21*, 102-107.

Weiss, F., Blum, G. S., & Gleberman, L. (1987). Anatomically Based Measurement of Facial Expressions in Simulated Versus Hypnotically Induced Affect. *Motivation and Emotion, 11*, 67-81.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*, 655-684.

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience, 18*, 411-418.

Wheelwright, S., Baron-Cohen, S., Goldenfeld, N., Delaney, J., Fine, D., Smith, R., Weil, L., & Wakabayashi, A. (2006). Predicting autism spectrum quotient (AQ) from the systemizing quotient-revised (SQ-R) and empathy quotient (EQ). *Brain Research, 1079*, 47-56.

Wicker, B., Fonlupt, P., Hubert, B., Tardif, C., Gepner, B., & Deruelle, C. (2008). Abnormal cerebral effective connectivity during explicit emotional processing in adults with autism spectrum disorder. *Social Cognitive and Affective Neuroscience*, *3*, 135-143.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. *Neuron, 40*, 655-664.

Williams, M. A., Moss, S. A., Bradshaw, J. L., & Mattingley, J. B. (2005). Look at me, I'm smiling: Visual search for threatening and nonthreatening facial expressions. *Visual Cognition*, *12*, 29-50.

Williams, J. H., Waiter, G. D., Gilchrist, A., Perrett, D. I., Murray, A. D., & Whiten, A. (2006). Neural mechanisms of imitation and 'mirror neuron' functioning in autistic spectrum disorder. *Neuropsychologia, 44*, 610–621.

Williams, J. H., Whiten, A., & Singh, T. (2004). A systematic review of action imitation in autistic spectrum disorder. *Journal of Autism and Developmental Disorders*. *34*, 285–299.

Williams, J. H., Whiten, A., Suddendorf, T., & Perrett, D. I., 2001. Imitation, mirror neurons and autism. *Neuroscience and Biobehavior*. Rev. *25*, 287–295.

Wimmer, H., & Perner, J. (1983). Beliefs About Beliefs - Representation and Constraining Function of Wrong Beliefs in Young Childrens Understanding of Deception. *Cognition, 13*, 103-128.

Wing, L. (1981). Sex-Ratios in Early-Childhood Autism and Related Conditions. *Psychiatry Research, 5*(2), 129-137.

Wing, L., Gould, J., Yeates, S. R., & Brierley, L. M. (1977). Symbolic play in severely mentally retarded and in autistic children. *Journal of Child Psychology and Psychiatry*, *18*, 167-178.

Wittgenstein, L. (1980). *Remarks on the Philosophy of Psychology*, vol. 2, G.H. von Wright and H. Nyman (eds.), C.G. Luckhardt and M.A.E. Aue (trans.), Oxford: Blackwell.

Wong, T. K. W., Fung, P. C. W., Chua, S. E., & McAlonan, G. M. (2008). Abnormal spatiotemporal processing of emotional facial expressions in childhood autism: dipole source analysis of event-related potentials. *European Journal of Neuroscience*, *28*, 407-416.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*, 1-34.

Woodward, A. L., & Summerville, J. A. (2000). Twelve-month-old infants interpret action in context. *Psychological Science*, *11*, 73-77.

World Health Organisation (WHO) *Tenth revision of the International Classification of Diseases and related health problems*. *Clinical Descriptions and diagnostic guidelines*. Geneva: WHO; 1992.

Yoshikawa, S., & Sato, W. (2008). Dynamic expressions of emotion induce representational momentum. *Cognitive, Affective*, *& behavioural Neuroscience*, *8*, 25-31.

Young, A. W., Aggleton, J. P., Hellawell, D. J., Johnson, M., Broks, P., & Hanley, J. R. (1995). Face Processing Impairments after Amygdalotomy. *Brain, 118*, 15-24.

Young, A. W., Ellis, A. W., Flude, B. M., Mcweeny, K. H., & Hay, D. C. (1986). Face Name Interference. *Journal of Experimental Psychology-Human Perception and Performance, 12*, 466-475.

Young, D. M., & Beier, E. G. (1977). Role of Applicant Nonverbal-Communication in Employment Interview. *Journal of Employment Counseling, 14*(4), 154-165.

Zaitchik, D. (1991). Is Only Seeing Really Believing - Sources of the True Belief in the False Belief Task. *Cognitive Development, 6*, 91-103.

Zalla, T., Labruyere, N., Clement, A., & Georgieff, N. (2010). Predicting ensuing actions in children and adolescents with autism spectrum disorders. *Experimental Brain Research, 201*, 809-819.

APPENDIX

Electromyographical (EMG) responses to Joy/Anger-to-neutral videos: a preliminary study

A1.1 Experiment 14

A1.1.1 Background

In the last decade several studies showed that the mere observation of facial expressions produces rapid and inconscious responses in the observer's facial muscles. For example, facial expressions of negative and positive affect produce rapid specific facial EMG responses over Corrugator Supercilii (the 'frowning' muscle) and the Zygomatic Major (the 'smiling' muscle), respectively (Dimberg & Thunberg 1998; Dimberg et al., 2000; Larsen et al., 2003). It has been suggested that these motor responses to the sight of facial expressions may reflect activity in mirror mechanisms enabling an immediate grasp of the emotional state of mind of mind of the agent through embodied simulation (Gallese et al., 2004).

In this thesis it was reported that Joy-to-neutral and Anger-to-neutral video-clips generate a perceptual distortion for the last neutral frame of the sequence, which is consistenly rated as slightly angry and slightly happy, respectively. A possible interpretation of this 'overshoot' effect is that Joy/Anger-to-neutral perceptual histories gave rise to 'emotional anticipation': i.e. the involuntary anticipation of how the emotional state of the other would develop in the immediate future (Chapter 2 and 3; Palumbo & Jellema, under review). It has also been suggested that a sudden change from a full-blown happy expression into a neutral expression may constitute a negative signal (i.e. threat), which would trigger an avoidance reaction in the observer. Similarly, a change from a full-blown angry expression into a neutral expression would be processed as a positive signal (i.e. reward/approach reaction (Mühlberger et al., 2011). It has been hypothesized that the Joy-to-neutral video sequence would lead the observer to involuntarily anticipate that the agent's immediate future emotional state will move into a negative affective state, and, similarly, that the Anger-to-neutral video sequence would lead the observer to involuntarily anticipate that the agent's future emotional state will move into a negative affective state (Jellema et al., 2011).

This interpretation was supported by the findings reported in Chapters 2 and 3, which suggested that TD observers reacted to the agent's emotional state rather than to the literal facial characteristics. This internal 'matching' reaction of the observer to the emotional/affective state of the actor could have possibly triggered an anticipatory mechanism enabling the observer to implicitly predict the agent's future emotional state. Such a mechanism would be vital for intersubjective understanding.

The role played by the motor experience of the observer in the perception of facial expressions is in line with theories of embodied cognition (Gallese, 2007). Thus, a simulation of the agents' dynamic facial expression activates (reflexively) motor schema, which run a 'simulation' of the observed behavior telling the observer what she/he would feel if she/he would have exhibited that behavior. This experience is then attributed to the agent, coloring the visual perception of the agent. An alternative model claims that the meaning and/or goal of the observed action is first deciphered in areas outside the mirror mechanism areas, and only then (after the goal computations have been performed) fed to the mirror mechanisms (Csibra, 2007). In this view, the mirror neuron activity reflects the anticipated action rather than the observed action. However, according to both these models (Csibra, 2007; Gallese, 2007), anticipation of the agent's emotional state of mind should be accompanied by activation in corresponding muscles.

We recorded EMG activity from TD individuals while they passively observed the video clips (Experiment 14). Specifically, we tested whether the EMG activity of the observer at the end of Joy-to-neutral video clips (i.e. video-clips lacking the typical features of an angry facial expression) would increase in the 'frowning' muscles (Corrugator Supercilii). Similarly, we tested whether EMG activity would increase in the 'smiling' muscle (Zygomaticus Major) at the end of Anger-to-neutral video-clips (i.e. video-clips lacking the typical feature of a happy facial expression). Thus, the aim of EMG experiment was to examine whether our previous behavioural findings are in line with an embodied simulation account of emotional anticipation. This experiment should, however, be considered as preliminary, because, due to frequent artifacts in EMG traces, not enough participants could be included in the analysis to allow drawing firm conclusions.

A1.1.2 Method

A1.1.2.1 Participants

The current pilot study was conducted at the Department of Neuroscience, Institute of Physiology, in the Faculty of Medicine and Surgery, University of Parma, Italy. Fifteen TD students of Italian nationality with normal or corrected-to-normal vision, naïve to the aim of the experiment, took part in the study on voluntary basis, and provided written consent prior to the experiment. After applying exclusion criteria (see data reduction below) eleven individuals (mean age = 29.5; SD = 4.2; 5 females and 6 males) were included in the final sample. The study received ethical approval by the Department of Neuroscience of Parma University.

A1.1.2.2 Stimuli

Joy-to-neutral and Anger-to-neutral videos similar to those in the control conditions of the experiments in Chapter 2 were employed (all clips ended in a neutral expression) Previous EMG studies with static pictures of happy and angry facial expressions showed that distinct facial reactions during the first second of exposure are most clear-cut during the period 500–1000 ms after stimulus onset (Dimberg & Thunberg, 1998). In the current study the duration of the first and last frames of the video-sequence was adjusted so as to capture maximal muscle activation. Thus, the first frame (100% happy or angry expression) lasted 700 ms, while the last frame (the neutral to-be-evaluated expression) remained on the screen for 1000 ms. The duration of the morphed sequence was 270 ms (9 x 30 ms), the same as in Experiment 1 (see Figure A1). The total duration of the stimuli presentation was 1970 ms.

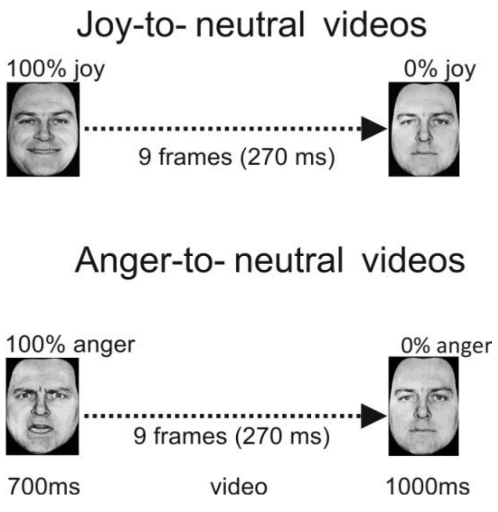

Figure A1. Schematic illustration of the stimulus presentations in Experiment 14. Joy-to-neutral videos started with a facial expression of intense joy (100%), which gradually morphed into a neutral expression (top panel). Anger-to-neutral videos started with a facial expression of intense anger (100%), which gradually morphed into a neutral expression (bottom panel).

A1.1.2.3 Experimental procedure

Participants were individually tested in a sound-attenuated room. All gave written consent for taking part in the study. In order to lower electrode impedance (< 10 kOhm), participants underwent a light scrubbing of the facial areas with alcohol. Bipolar Ag/AgCl (.25 cm diameter) electrodes were first covered with a layer of electrolytic gel and then placed over the Corrugator Supercilii and Zygomaticus Major in the direction of the muscle's fiber orientation on both sides (left and right) of the face (see Figure A2). The inter-electrode spacing was approximately 1 cm. The ground electrode was placed over the forehead. Correct placement of the electrodes was confirmed by checking the EMG in response to deliberate smiles and frowns produced by the participant.

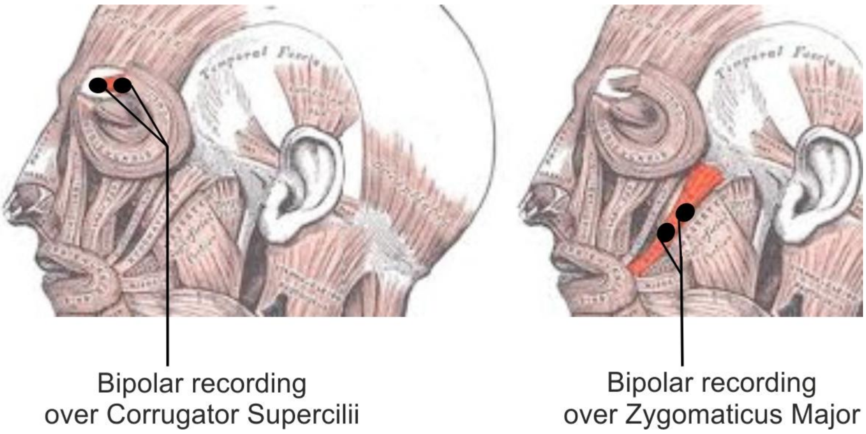

Figure A2. Illustration of the recording positions of the surface electrodes over the Corrugator Supercilii and the Zygomaticus Major.

Participants were seated at a viewing distance of 100 cm from a PC screen (21-inch monitor, 1600 x 1200 pixels, 76 Hz). The stimuli were presented using E-Prime (v. 1.2; Psychology Software Tools, Inc.).

The experiment started with a calibration phase, in which participants rated the static neutral expression of each actor presented for 600 ms using the same 5-point scale (ranging from slightly angry, via neutral, to slightly happy) as in Experiment 1. After the calibration participants were presented with 8 practice trials, followed by 64 experimental trials randomly intermixed (8 actors x 2 perceptual histories x 4 repetitions). Each trial started with a fixation cross presented for 700 ms, followed by the video-clip. As soon as the video-clip ended, participants were prompted to evaluate the last neutral frame of the sequence on the same 5-point scale as used in the calibration phase. Further procedures and apparatus were as in Experiment 1. EMG signals were recorded continuously during the experimental trials and amplified 1000 times (CED amplifiers: Cambridge Electronic Device, Cambridge, UK). The sampling rate was set at a frequency of 2.5 kHz. Online low-pass filtering was applied with a cut-off at 500Hz in order to eliminate aliasing artifacts before A/D conversion. A 50-Hz notch filter was also applied to reduce environmental electrical noise.

A1.1.2.4 Data reduction and analysis

Behavioural data

Trials in which the RTs fell below 250 ms or above 3000 ms were considered outliers and were removed (3.2%). Participants were excluded if more than 25% of the RT values were outside the 250 – 3000 ms range (n = 1). The ±2.5 SD rule applied to the difference between the mean ratings for the Joy-to-neutral and Anger-to-neutral videos did not exclude any participants. However, three more participants were excluded because they pressed the key corresponding to ‘slightly angry’ on the scale (1 = slightly angry) on more than 90% of the total trials.

The calibration scores were used to adjust the scores in the experimental trials. That is, a calibration factor of [3.00 minus the calibration score] was added to the raw scores. For illustrative purposes the calibration scores for the neutral expression of each actor are shown (n = 11) in Figure A3. The calibration procedure allowed performing one-sample t tests with test value 3.00 for each perceptual history. All statistical analyses were performed on the calibrated scores.

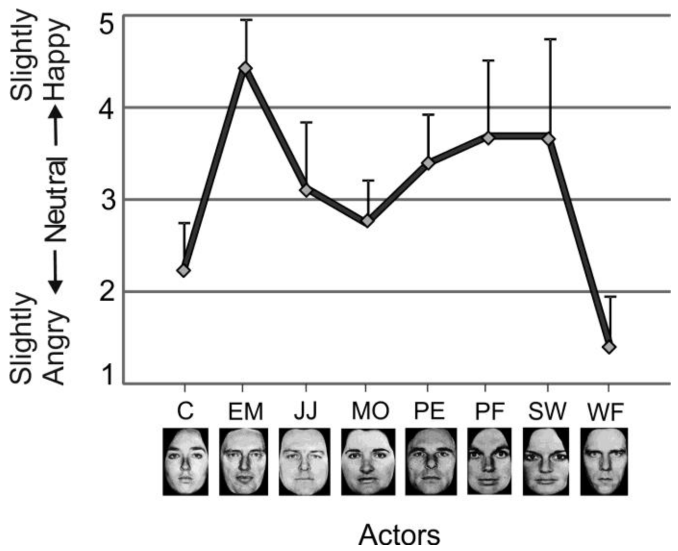

Figure A3. Calibration scores in Experiment 14. Ratings on the 5-point scale (vertical axis) for the neutral expressions of the 8 actors presented in the calibration phase. Error bars indicate +1SD.

EMG data

EMG data were processed offline applying a high-pass filter with a cut-off at 30Hz, and were fully-rectified (Spike 2; Cambridge Electronic Design Ltd, Cambridge, UK). The EMG activity was also inspected for eventual artefacts occurring during the fixation cross (700 ms), the first 100% happy or angry frame (700 ms) and the last neutral frame (1000 ms). Artifacts were removed manually (Signal 3; Cambridge Electronic Design Ltd, Cambridge, UK).

Further computations and data analysis were performed in MATLAB 5.3 (1994-2012, The MathWorks, *Inc*). First, trials which were rejected according to the RTs cut-off (3.2%; see section A1.2.4.1) were also excluded from the EMG data analysis. Three periods of interest were identified: the fixation cross, which was used as baseline (700 ms; phase 1), the first 100% happy or angry frame (700 ms; phase 2) and the last neutral frame (1000 ms; phase 3). Only the last 500 ms of each phase were taken into account, resulting in three 500 ms bins: 'bin 1' (baseline), 'bin 2' (100% happy or angry first frame) and 'bin 3' (last neutral frame), as illustrated in Figure A4. In this explorative study the statistical analyses were performed on a subset of 9 participants from the total sample of 15 participants. For the 6 excluded participants the EMG was too noisy and contained too many artifacts. For the Joy-to-neutral videos, it was predicted to find an increase of the activity in the left and right Corrugator in bin 3 as compared to bin 2. For the Anger-to-neutral videos, it was predicted to find an increase of the activity in the left and right Zygomaticus Major in bin 3 as compared to bin 2.

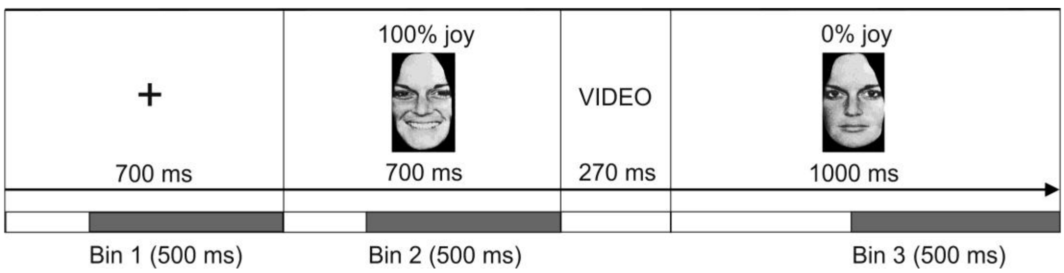

Figure A4. Illustration of the presentation sequence with the three 500 ms bins corresponding to the baseline (bin 1), first frame 100% emotion (bin 2) and last neutral frame (bin 3).

A1.1.3 Results

A1.1.3.1 Behavioural data

The results are shown in Figure A5. The typical overshoot response bias was found: the last neutral frame was perceived as more angry in the joy-videos (M = 2.70, SD = .31) as compared to the anger-videos (M = 3.37, SD = .29). The mean ratings in the joy-videos and anger-videos differed significantly from each other ($t(10) = -10.3$, $p < .0001$) and each differed significantly from 3.00 (joy-video: $t(10) = -3.2$, $p = .009$; anger-video: $t(10) = 4.19$, $p = .002$; two-tailed).

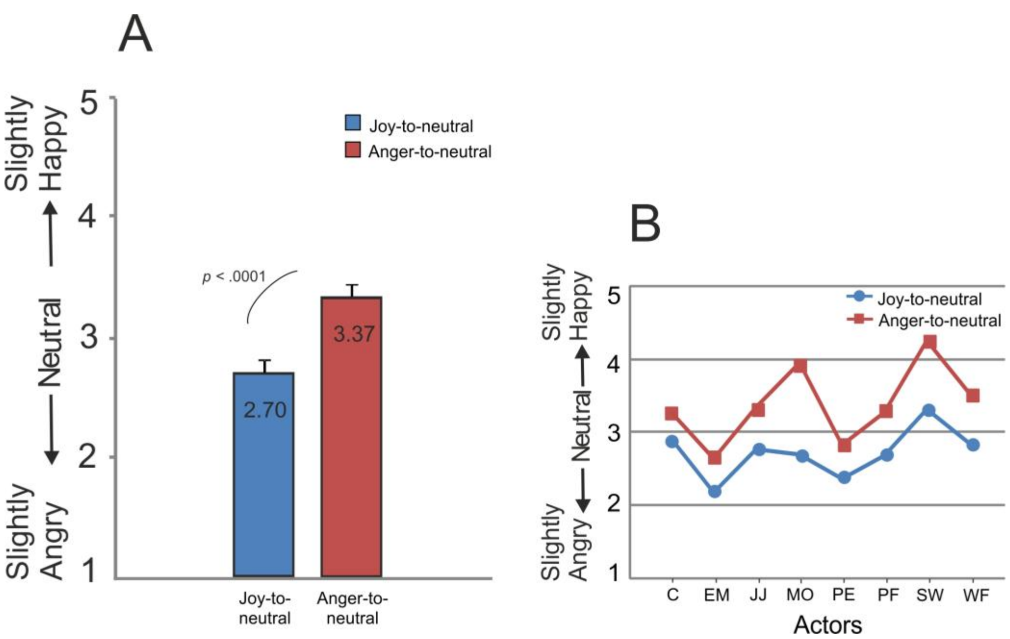

Figure A5. Results for the behavioural task of Experiment 14. (A) Scores for the neutral expressions at the end of the Joy-to-neutral and Anger-to-neutral sequences. Error bars indicate SEM. (B) Ratings for each of the eight actors with Joy-to-neutral and Anger-to-neutral videos are shown to illustrate the consistency across actors.

A.1.1.3.2 EMG data

The results are shown in Figure A6. Paired-samples t-tests in the Joy-to-neutral condition reported a trend for an increase in activity in the right Corrugator for the last neutral expression (bin 3; M = .51, SD = .63) relative to the maximally happy expression (bin 2; M = -.12, SD = .64) (t(8) = 1.96, p = .086; two-tailed). In the left Corrugator, there was no difference in activity between bins 2 and 3 (t(8) = 1.35, p = .214; two-tailed).

In the Anger-to-neutral condition, paired sample t-tests showed significant increases in activity in both the left Zygomaticus in bin 3 (M = .80, SD = .31) compared to bin 2 (M = .00, SD = .77) t(8) = 3.55, p = .008 and in the right Zygomaticus (bin 3, M = .42, SD = .33 ; bin 2, M = -.18, SD = .35; t(8) = 5.07, p = .001; two-tailed).

Additional one sample t-tests also revealed that the Corrugator activity during the maximally happy frame (bin 2) and the Zygomaticus activity in the maximally angry frame (bin 2), did not significantly differ from the activity during the baseline (bin 1) (left Corrugator: t(8) = .79, p = .451; right Corrugator: t(8) = .55, p = .599; left Zygomaticus: t(8) = .02, p = .983; right Zygomaticus: t(8) = 1.54, p = .163; two-tailed). This result excludes the possibility that the increase in activity observed in the Corrugator during the last neutral frame of the videos (bin 3) as compared to bin2, in the Joy-to-neutral condition, actually reflected a rebounce from an inhibition, i.e. an inhibition in response to the maximally happy expression (bin 2).

Similarly, the result excludes the possibility that the increase in activity observed in the Zygomaticus during the last neutral frame of the videos (bin 3) as compared to bin2, in the Anger-to-neutral condition, actually reflected a rebounce from inhibition in response to the maximally angry expression (bin 2).

Finally, one sample t-tests in the Joy-to-neutral condition showed a trend for an increase in activity in the right Corrugator for the last neutral expression (bin 3) relative to the baseline (bin 1) ($t(8) = 2.46$, $p = .039$). In the left Corrugator there was no difference in the activity between bin 3 and bin 1 $t(8) = -.06$, $p = .954$; two-tailed).

In the Anger-to-neutral condition, one sample t-tests revealed a significant increase in activity in both the left and the right Zygomaticus in bin 3 compared to bin 1 (left Zygomaticus: $t(8) = 7.67$, $p < .0001$; right Zygomaticus: $t(8) = 3.81$, $p = .005$; two-tailed).

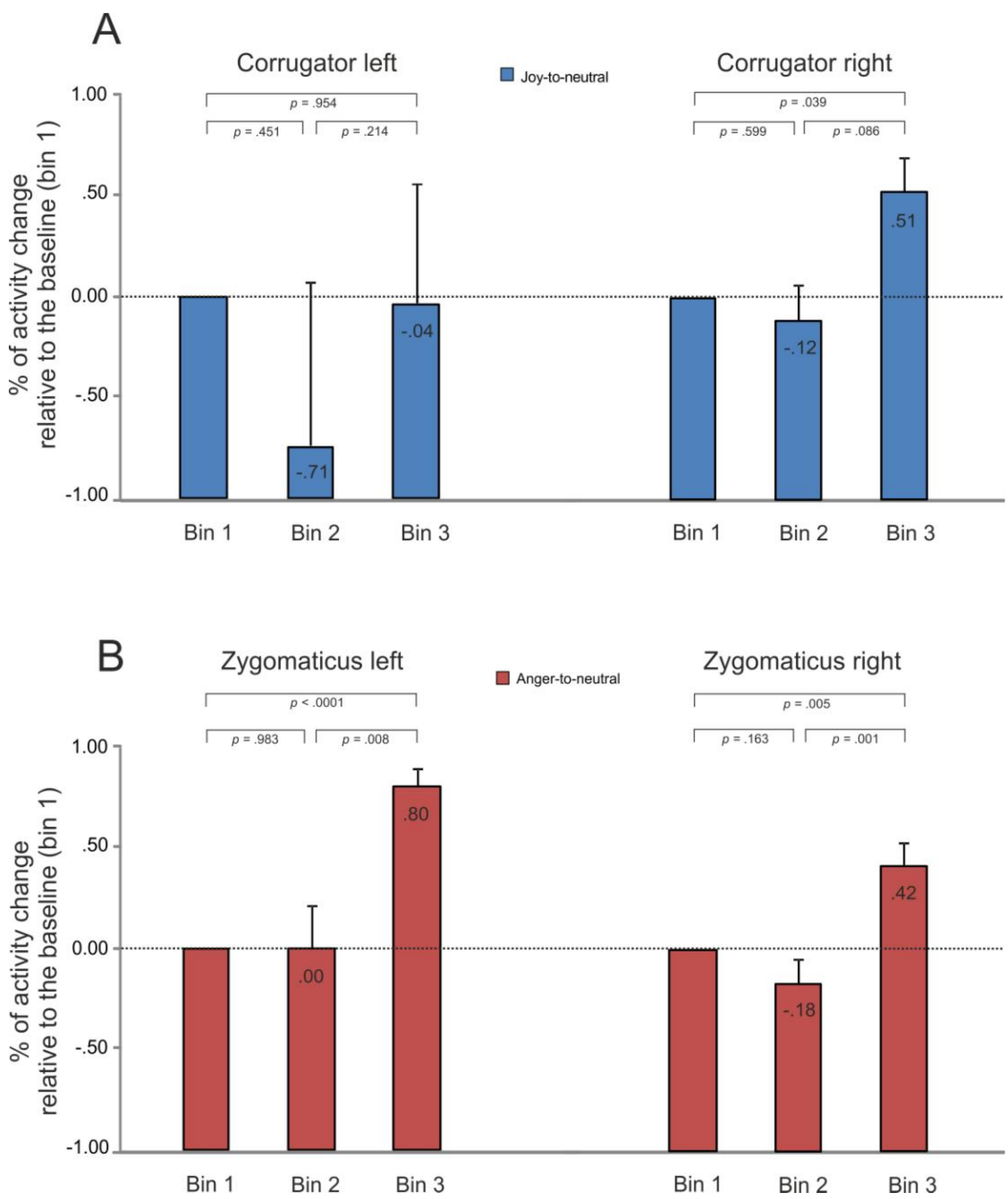

Figure A6. Results for the EMG activity of Experiment 14. (A) Increase of activity in the Corrugator Supercilii during the last neutral expressions (bin 3) of the Joy-to-neutral sequence (x axis). (B) Increase of activity in the Zygomaticus Major during the last neutral expression (bin 3) of the Anger-to-neutral sequence (x axis). The EMG activity is expressed as percentage of bin 1 (y axis). Error bars indicate SEM.

A1.1.4 Discussion

The main purpose of employing the EMG technique in combination with the emotional anticipation task was to investigate the role played by embodied mechanisms of motor resonance in understanding and anticipating others' emotions from dynamic facial displays. Therefore, the specific objective of this study was to test whether the perceptual distortion, as found for the neutral expression at the end of Joy/Anger-to-neutral videos, would be associated with a congruent response in the observer's facial reactions.

It was found that the overshoot response bias in the Joy-to-neutral condition was associated with an increased activity in the right Corrugator, while the overshoot bias in the Anger-to-neutral condition came along with an increased activity in the right and left Zygomaticus. Overall, at least in this selected sample of 9 (out of 15) participants, the observed trend seems to be in line with embodied simulation models, both the 'direct matching' model of action/emotion mirroring (Gallese, 2007; Rizzolatti et al., 2001) and the 'reconstructive matching' model (Csibra, 2007). However, as the current data represent a selected, small, sample of participants, these results should be taken with caution, and as yet no firm conclusions can be drawn from it.

This pilot study contained several methodological limitations. Some modifications of the paradigm may improve the clearness of the EMG signal. For instance, the two types of video-clips (Joy and Anger) could be presented in two separated blocks. This might prevent possible interferences with ongoing activity, especially because the speed of the video-sequences was relatively high. At the moment almost all EMG studies using facial expressions adopted static presentations lasting several seconds (Dimberg, 1982). Moreover, it would be desirable to introduce an intertrial interval of several seconds. This would facilitate the relaxation of the muscle's fibers before starting with a new trial, which may improve the reliability of the baseline.

These preliminary findings seem to be a starting point for further investigations. The study needs to be replicated after applying the above methodological improvements on a much larger sample of participants. Further, it would be interesting to employ the EMG technique also with 'fake' Joy/Anger-to-neutral videos (see Chapter 4). One would expect to find smaller muscle activation when the observer is presented with 'fake' Joy/Anger-to-neutral sequences as compared to videos of genuine expressions. If this prediction would be confirmed then it would not only help to clarify the results obtained so far with the 'fake' manipulation, but it may also suggest that embodied simulation based on 'direct matching' mirroring is an integral part of the emphatic process enabling our immediate understanding of others' facial expressions.